

Evaluation of Transformer and BRNN Models for Machine Translation on Medical Text Corpora

Daniel Kayl

LILY Workshop Fall 2019

Dec. 13, 2019

CPSC 490 Project Proposal

- Originally aimed at evaluating different translation models using OpenNMT
- Wanted to use BERT-based models to try for improvement

Two Datasets: UFAL Medical Corpus and MeSpEn IBECS

UFAL Medical Corpus	EN-Train	EN-Test	ES-Train	ES-Test
Vocab Size	131066	48244	149209	56274
Longest Sent Len	458	245	444	275
Shortest Sent Len	1	1	1	1
Median Sent Len	8	8	9	9

Table 2: Statistics for UFAL Medical Corpus data.

MeSpEn IBECS Corpus	EN-Train	EN-Test	ES-Train	ES-Test
Vocab Size	274134	69397	278948	75529
Longest Sent Len	227	251	279	332
Shortest Sent Len	1	1	1	1
Median Sent Len	19	19	22	22

Table 3: Statistics for MeSpEn IBECS Corpus.

Sample Documents

- MeSpEn:
 - “A new trend is to be seen in the drug use profile over the last few years, with a fall in heroin use and a rise in that of cocaine. However, a new use has appeared, simultaneously taking heroin and cocaine, known as a speedball. Clinically, there are clear similarities between speedball users and heroin users and both differ from those taking cocaine. This study endeavours to ascertain if the emotional response is also influenced by the kind of substance used and if said emotional response in those taking speedball is more similar to that of cocaine users or that of heroin users. We studied 75 subjects (34 using cocaine, 11 heroin and 30 speedball), using the Instrumento Cl nico de Evaluaci3n de la Respuesta Emocional (ICERE), which is based on the emotional response model by Peter J. Lang. The ICERE comprises 25 images selected from the International Affective Picture System (IAPS) and the response was measured using the Self Assessment Maniky (SAM). The results show that differences exist in the response of drug dependent individuals to emotional stimuli in respect of their preferred substance use, no differences exist between speedball and heroin users, and differences do exist between both these groups and cocaine users”

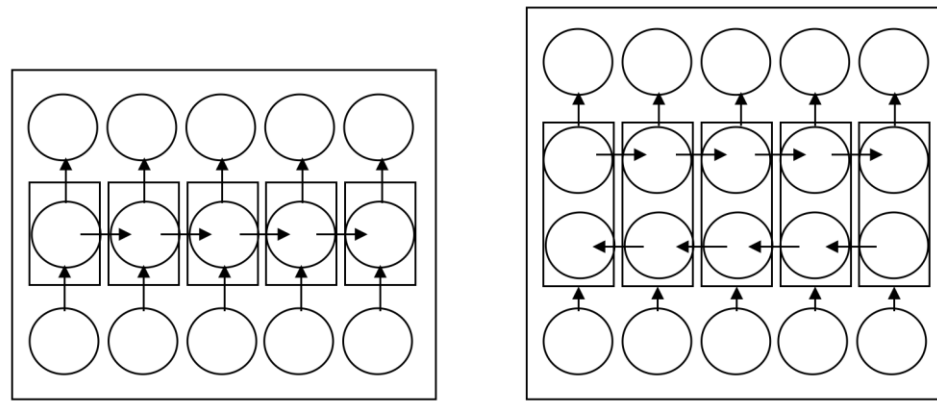
Sample Documents Cont.

- UFAL

- Subtitles: Still, he said, "it is unfortunate that almost three years pass from when a parent first expresses concerns to his or her pediatrician and when the child gets a diagnosis of an autism spectrum disorder."
- Bitch!
- In clinical pharmacokinetic studies of tipranavir/ ritonavir in combination with omeprazole (40 mg qd), no clinically important changes in tipranavir/ ritonavir plasma concentrations were observed and thus no tipranavir/ ritonavir dose adjustment is required.
- Novo Nordisk A/ S Novo Allé DK-2880 Bagsværd Denmark

OpenNMT Models

- Used OpenNMT, an open-source machine translation toolset
- Transformer model based off of Zhang et al. (2018)
- BRNN



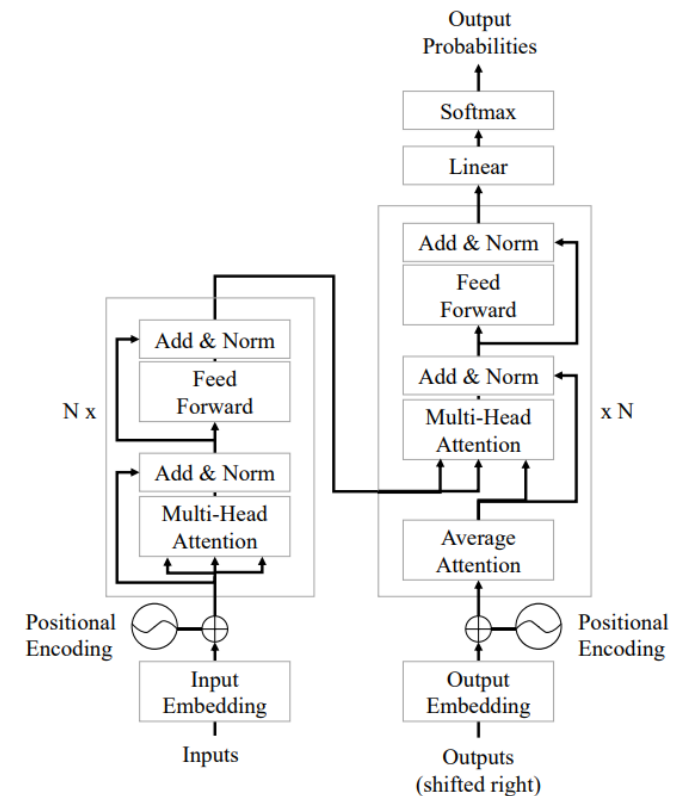
(a)

(b)

Structure overview

(a) unidirectional RNN

(b) bidirectional RNN



Google Translate

- Used Google Translate functionality in Google Sheets to compare OpenNMT results against general domain

40	Introduction. After systemic administration, recombinant adenoviruses are rapidly concentrated i	Loading...
41	At birth, vernix caseosa can cover the whole body surface or accumulate only on the back and in	Loading...
42	Objectives: Evaluation of the impact of EUS in clinical practice. Methods: All exploration performe	Loading...
43	Backgrounds: Improper use of incontinence absorbent products in adults can negatively affect se	Loading...
44	This study aimed at determining the relationship between various levels of direct and indirect ex	Loading...
45	Autism spectrum disorder (ASD) is a term used to describe a heterogeneous group of children wh	Loading...
46	Objective. To examine the main clinical and laboratory data of patients initially diagnosed with pc	Loading...
47	The aim of this paper is to reflect on supervision, an essential instrument for professional develo	Loading...
48	Workplace safety interventions, designed for either research or professional practice purposes, s	Loading...
49	Introduction. The internal hernia occurs rarely, especially in children, with picture of acute abdon	Loading...
50	Background and objective: The prevalence of extended-spectrum beta-lactamase (ESBL) producir	Loading...
51	The location of plaque-associated gingivitis at the gingival portion of the tooth plays an essential	Loading...
52	Introduction: women may be nutritionally vulnerable across life stage and due to different physic	Loading...
53	First prize in the VII edition of the award winning work Betadine Â® for nursing 2012. The dermat	Loading...
54	When bariatric and metabolic surgery initially began in Spain, it was a subject of debate, due to n	Loading...
55	Introduction: Few studies have evaluated the efficacy and reliability of weight loss-focussed prep	Loading...

```
function translate() {
  var ss=SpreadsheetApp.getActiveSpreadsheet();
  for (var sheetnum=0; sheetnum < ss.getSheets().length; sheetnum++) {
    var sheet=ss.getSheets()[sheetnum];
    var selection=sheet.getDataRange();
    var columns=selection.getNumColumns();
    var rows=selection.getNumRows();
    for (var row=1; row <= rows; row++) {
      //gets rid of cell-initial =
      //   var col1=selection.getCell(row, 1);
      //   if(col1.getFormula()[0] == '=') {
      //     col1.setValue(col1.getFormula().slice(1));
      //   }
      var cell=selection.getCell(row, 2);
      cell.setValue(Utilities.formatString("=GOOGLETRANSLATE(A%d, \"en\", \"es\")", row));
    }
  }
}
```

Results

- BLEU Scoring
 - Uses human translations as gold standard

Corpus-Model	BLEU-1	BLEU-2	BLEU-3
MeSpEn-Google	0.1288	0.0131	0.0034
MeSpEn-OpenNMT(transformer_UFAL)	0.0562	0.0028	0.0004
MeSpEn-OpenNMT(brnn_UFAL)	0.0558	0.0030	0.0004
MeSpEn-OpenNMT(transformer_mespen)	0.0567	0.0029	0.0003
UFAL(testing)-Google	0.1298	0.0122	0.0028
UFAL(testing)-OpenNMT(transformer)	0.1301	0.0046	0.0005
UFAL(testing)-OpenNMT(brnn)	0.1311	0.0046	0.0004

Table 4: BLEU scores for all models tested.

Problems Encountered

- Buggy OpenNMT code
- Messy datasets
- PyTorch really hampered translations...

Future Directions

- Using BERT-based models
- Clean dataset better
- More data for training (rest of MeSpEn corpora)