CUTIE_Learning_to_Understand_Documents

Nisim Hurst

Tuesday 26 May 2020

CUTIE: Learning to Understand Documents

The article was written by (**zhao2019cutie**). It was was cited 4 times according to Google Scholar. The task performed was semantic instance segmentation over a text grid using convolutional neural networks at character level. They used the average precision metric (AP) in two forms, one strict and the other soft, over 9 classes for each field they tried to detect.

Hypothesis

By taking into account both position and semantic meaning for segmenting 2D documents, a CNN will be able to outperform other algorithms based on named entity recognition.

Evidence and Results

Dataset

- Self-built dataset with 4,484 labelled receipts. 3,375 samples are used for training and 1,125 for testing, with around 1500 samples for each of the three document types (Taxi, Meals/Entretainment and Hotel).
- ICDAR 2019 task 3 with 1000 scanned receipt images. 627 for training/validation, 157 validation, 470 training. 55 for testing.
- 8 key information classes and one don't care class.
  - DontCare
  - VendorName
  - VendorTaxID
  - InvoiceDate
  - InvoiceNumber
  - ExpenseAmount
  - BaseAmount
  - TaxAmount
  - TaxRate

Architecture

layer name	operations	input dimension	output dimension	comments
embedding layer	-	20000	128	
conv block	$[3 \times 5] \times 4$	256	256	stride=1
atrous conv bloc	$[3 \times 5] \times 4$	256	256	stride=1, rate=2
ASPP module	$[3 \times 5] \times 3$, global pooling, concat, 1×1	256	256	stride=1, rate={4,8,16}
shorcut layer	concat, 1×1	256	64	
output layer	1×1	64	9	

Results

- CloudScan (**palm2017cloudscan**)
- BERT (**devlin2018bert**)
- CUTIE-A
- CUTIE-B

Contribution

- A dataset of roughly 4500 grid texts with labels for their corresponding 9 classes.
- Convolutional architecture using multiple subnetworks with high-resolution.
- Convolutional architecture using the atrous convolution and a single network.

Weaknesses

The authors don't tokenize each word into a meaningful entity. Instead, they use a predefined dictionary.

Future Work