Unconstrained Product Categorization with Sequence-to-Sequence Model

Maggie Yundi Li[1], *Liling Tan*[3], *Stanley Kok*[2] *and Ewa Szymanska*[3]
National University of Singapore[1,2] and Rakuten Institute of Technology (Singapore)[3]
a0131278@comp.nus.edu.sg[1], skok@comp.nus.edu.sg[2], {first.last}@rakuten.com[2]

Introduction

Sample of Training Data (Product Titles):

Category: 3292>1041>4175>4258
Canon EOS M10 Mirrorless Digital Camera with 15-45mm Lens + 16GB Memory Card + Camera Case
Canon 6163B001M PowerShot ELPH 530HS White 10.1MP
Panasonic Lumix DMC-GF7 Mirrorless Micro Four Thirds Digital Camera (Black Body Only)
Category: 3292>1041>4380>4953
Canon PowerShot Elph 360 HS Wi-Fi Camera + 32GB + Case + Battery + Selfie Stick + Sling Strap + Kit
Fujifilm X-E3 4K Digital Camera & 23mm f/2 XF Lens (Silver)
Category: 3292>1041>4380>4374
Canon EF 70-200mm f/2.8L IS II USM Telephoto Zoom Lens Deluxe Accessory Bundle

Input: '*Canon 9167b001 12.8 Megapixel Powershot(R) G1 X Mark Ii Digital Camera*'

Prediction of Hierarchical Categories:

3292>1041>4380>4258

(Does not exist in training labels)

Data Characteristics

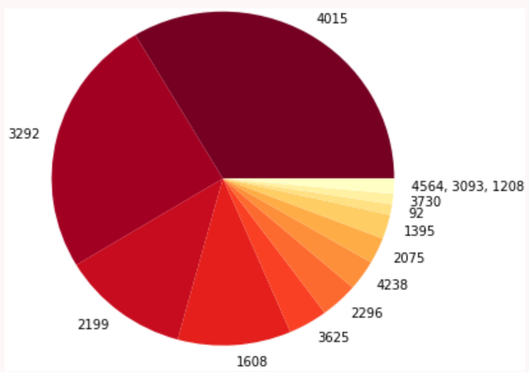


Figure 1: Distribution of First-Level Categories

Data Characteristics

Class imbalance

- 3000+ unique category labels
- 8.7% of 800K titles = 2199>4592>12

Noisy Product Titles

- Systematic noise from non-ASCII characters

Model

Model: Single layer Seq2Seq with Attention[1]

Hyperparam:

```
{'RNN Cell':'GRU', 'Vocab Size':120000,
'Embed Dim':512, 'RNN Dim':1024,
'Dropout':0.2, 'Optimizer':'Adam',
'Batch Size':5000, 'Learning Rate':0.0001
'Beam Size':6}
```

Tool: Marian[2] (commit f429d4a)

Stop when: Perplexity falls below 1.1

Results

Our Model	Random Seed	Epoch	Cross-entropy	Perplexity
M1	0	77	0.8446	1.1835
M2	1	189	0.0191	1.0038
M3	1	470	0.0723	1.0145
M4	2	54	0.0542	1.0108

Table 1: Cross-entropy and Perplexity during Training

Phase	Model(s)	P	R	F
1	M1 (Baseline)	0.82	0.81	0.81
	M1-3	0.83	**0.83**	0.82
	M1-4	**0.8311**	0.8296	**0.8245**
2	M1-4	0.8267	0.8305	0.8256
	Best system (mcskinner)	**0.8697**	**0.8418**	**0.8513**

Table 2: Precision, Recall, F1 Scores on Held-out Test Set

Best Setup: Ensemble of 4 models (M1 – 4)

Analysis: Machine Created Category

Seq2Seq cross-pollinated categories; created 4238>4960>1286

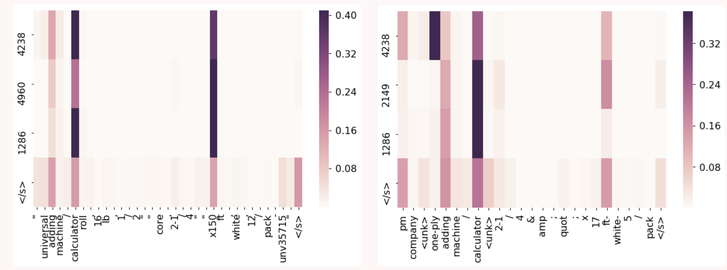
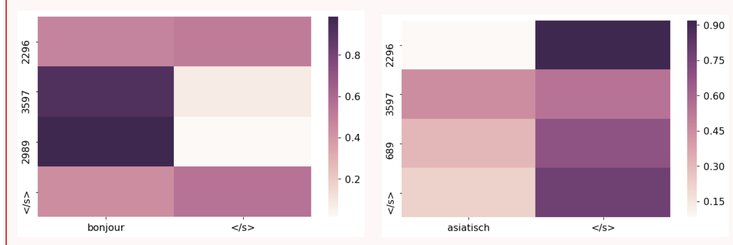


Figure 2: Attention Weights Alignments

(Left) *PM Company 07622 One-Ply Adding Machine/Calculator Rolls-2-1/4" x 17 ft- White- 5/Pack*

(Right) *"Universal Adding Machine/Calculator Roll, 16 lb, 1/2"" Core, 2-1/4"" x 150 ft,White, 100/CT - UNV35710"*

Analysis: Music Category



Model learns length information but fails on fine-grained category

Conclusion

- Seq2Seq model generates unconstrained labels
- Attention makes model interpretable
- Competitive results: 0.82 F-score (ranked 6th)

References:
[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
[2] Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, Andre F. T. Martins, and Alexandra Birch. 2018. Marian: Fast Neural Machine Translation in C++. In ACL 2018, System Demo.