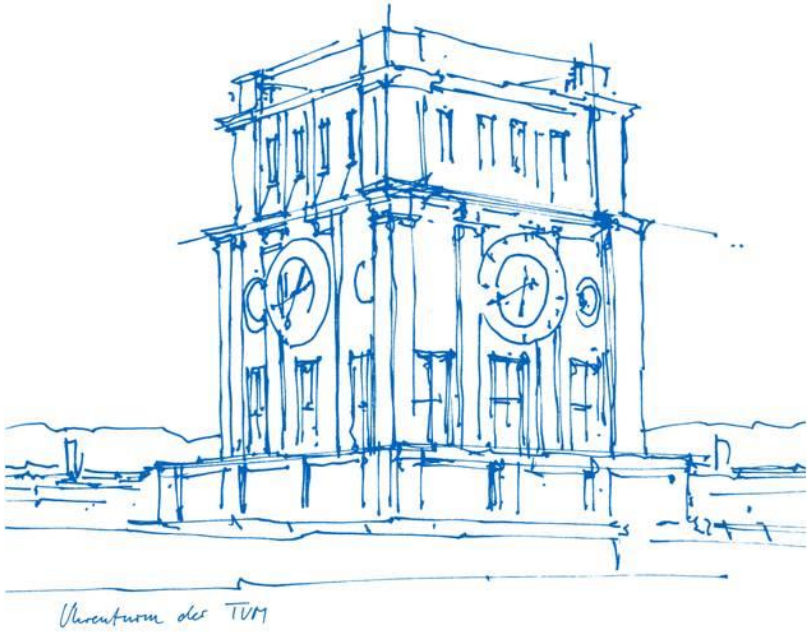
Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

Team 06

Michael Sorg

08.05.2019



Task

- Train a **category prediction model** on the Amazon product review dataset **based on XLING embeddings** per review
- Evaluate on German without training on German data
- Fine-tune and evaluate on our Organic Dataset for relevance, entity, and attribute classification

Word embeddings

BERT

„BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”

- Birectional transformer network
- Multilingual version exists

XLING

„Learning cross-lingual sentence representations via a multi-task dual-encoder model“

- Compute dense word vectors (512 dimensions) from sentences
- Embeddings can be fine-tuned (transfer learning)
- Multiple languages

Datasets

- **Kaggle Amazon Fine Food**

(568454, 10)

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...

- **Amazon Review Multilingual**

(678993, 15)

	marketplace	customer_id	review_id	product_id	product_parent	product_title	product_category	star_rating	helpful_votes	total_votes	vine	verified_purchase	review_headline	review_body	review_date
0	DE	10133	RVOG49N0H1FB6	B004TACMZ8	569741360	Bosch GMS120 Ortungsgerät digital multi-Scanner	Home Improvement	5	0	0	N	Y	Super	Delivery took a little bit more then i expecte...	2014-08-01
1	DE	19612	RNCMD6OLTP4HM	1846071224	785505948	The Wheels On The Bus: Favourite Nursery Rhyme...	Books	5	1	1	N	Y	Great compilation	We enjoy listening to the song as preparation ...	2014-12-04

- **Organic dataset**

(8823, 12)

	Author_ID	Author_name	Comment_number	Sentence_number	Domain_Relevance	Sentiment	Entity	Attribute	Sentence	Source_file	Annotator	Aspect
0	Justin-Ma	Justin Ma	521	1	0	NaN	NaN	NaN	Thanks for the thoughtful response.	quora.json	sumit	nan-nan
1	Justin-Ma	Justin Ma	521	2	0	NaN	NaN	NaN	I think we actually have a lot of common groun...	quora.json	sumit	nan-nan
2	Justin-Ma	Justin Ma	521	3	0	NaN	NaN	NaN	All I want to emphasize are my main points: Pr...	quora.json	sumit	nan-nan
3	Justin-Ma	Justin Ma	521	4	9	p	cg	pp	Industrialization is everything about producti...	quora.json	sumit	cg-pp

Questions

- Predict which categories / features ?
- Filter out English reviews in the German dataset?

- Also train XLING embeddings or leave them untouched?

To-do until 22.05

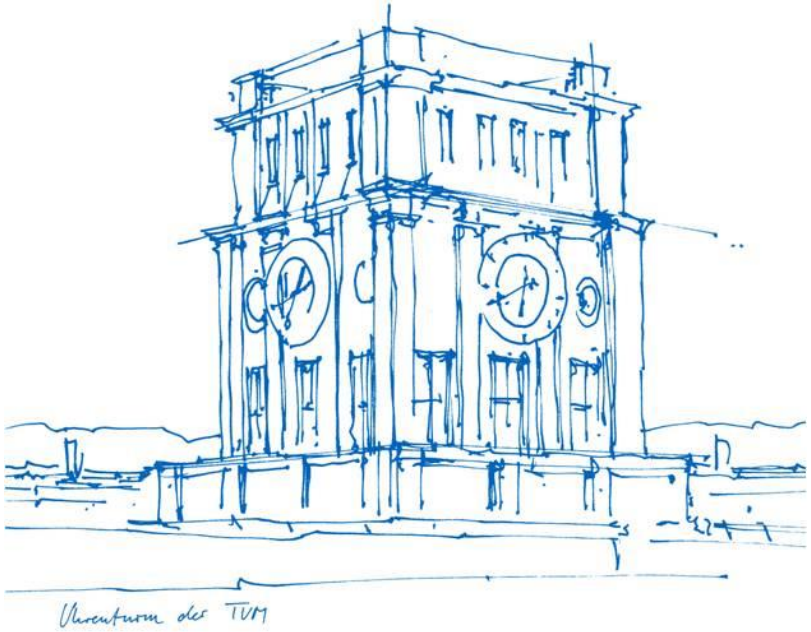
- Implement training and evaluation pipeline

Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

Team 06

Michael Sorg

22.05.2019



Amazon Multilingual Dataset

(678993, 15)

	marketplace	customer_id	review_id	product_id	product_parent	product_title	product_category	star_rating	helpful_votes	total_votes	vine	verified_purchase	review_headline	review_body	review_date
0	DE	10133	RVOG49N0H1FB6	B004TACMZ8	569741360	Bosch GMS120 Ortungsgerät digital multi-Scanner	Home Improvement	5	0	0	N	Y	Super	Delivery took a little bit more then i expecte...	2014-08-01
1	DE	19612	RNCMD6OLTP4HM	1846071224	785505948	The Wheels On The Bus: Favourite Nursery Rhyme...	Books	5	1	1	N	Y	Great compilation	We enjoy listening to the song as preparation ...	2014-12-04
2	DE	19612	R4AUOBI8YC0R8	0375851569	516548029	Dr. Seuss's Beginner Book Collection	Books	5	0	0	N	Y	Great Collection	Very great compilation. Interesting story and ...	2014-12-04
3	DE	19677	R1VSHIJ1RHIBTE	B0060SVG54	302116447	Zwei an einem Tag	Video DVD	5	0	0	N	Y	Guter Verfilmung	Den Film habe ich bereits vor lesen des Buches...	2015-07-16
4	DE	19999	R3JBLVALWSLCZD	B00EYQ6CVC	368843515	Dr. House - Die komplette Serie, Season 1-8 (L...	Video DVD	5	9	14	N	Y	Kauft diese Box!	Die Box ist super verarbeitet, sieht gut aus b...	2014-02-08

```
ar: 1
fa: 1
zh-cn: 1
bg: 2
ko: 2
lv: 2
tr: 2
cs: 4
lt: 5
ro: 8
vi: 8
tl: 15
fi: 17
ru: 17
sl: 18
hu: 21
pl: 21
sk: 23
hr: 24
id: 24
et: 32
sq: 33
pt: 54
no: 81
sv: 81
so: 82
da: 105
cy: 106
af: 108
it: 150
ca: 158
nl: 179
es: 180
fr: 307
en: 48660
de: 628461
```

Review

- For each dataset create tensorflow dataset loader (using dataset api)
- Amazon Reviews dataset does not fit on disk and into ram → change runtime in colab
- Todo: look into TFRecord

```
[75]  l = ["cache/reviews_Arts_Crafts_and_Sewing_5.json.gz",
           "cache/reviews_Baby_Products_5.json.gz"]
```

```
[86]  d = tf.data.TextLineDataset(l, compression_type="GZIP")
      d = d.shuffle(1024).batch(16).prefetch(16)
```

```
iterator = d.make_one_shot_iterator()
next_element = iterator.get_next()

with tf.Session() as sess:
    for i in range(6):
      value = sess.run(next_element)
      print(value)
```

```
[b'{"reviewerID": "A3TYDOH5JLRD37", "asin": "B000BNLLHW",
 b'{"reviewerID": "A1LV9A437V9X6K", "asin": "B000OMZXGU",
 b'{"reviewerID": "A2QGLMPBVZ3OYY", "asin": "B000AM7YJI",
 b'{"reviewerID": "A35G9GHVA9WHD4", "asin": "B0007XMDIM",
```

Review

☑ Full training pipeline with dummy architecture

Roadmap

1. network architecture search (train/test on amazon multilingual en/de)
   - w/o training xling embeddings

2. Additionally use full amazon reviews dataset

3. Test and fine-tune on organic dataset

Questions

- Docker instances → Google account required

(568454, 10)

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...

De: 600k, UK: 1,7m, US: 7M

(678993, 15)

	marketplace	customer_id	review_id	product_id	product_parent	product_title	product_category	star_rating	helpful_votes	total_votes	vine	verified_purchase	review_headline	review_body	review_date
0	DE	10133	RVOG49N0H1FB6	B004TACMZ8	569741360	Bosch GMS120 Ortungsgerät digital multi-Scanner	Home Improvement	5	0	0	N	Y	Super	Delivery took a little bit more then i expecte...	2014-08-01
1	DE	19612	RNCMD6OLTP4HM	1846071224	785505948	The Wheels On The Bus: Favourite Nursery Rhyme...	Books	5	1	1	N	Y	Great compilation	We enjoy listening to the song as preparation ...	2014-12-04

(8823, 12)

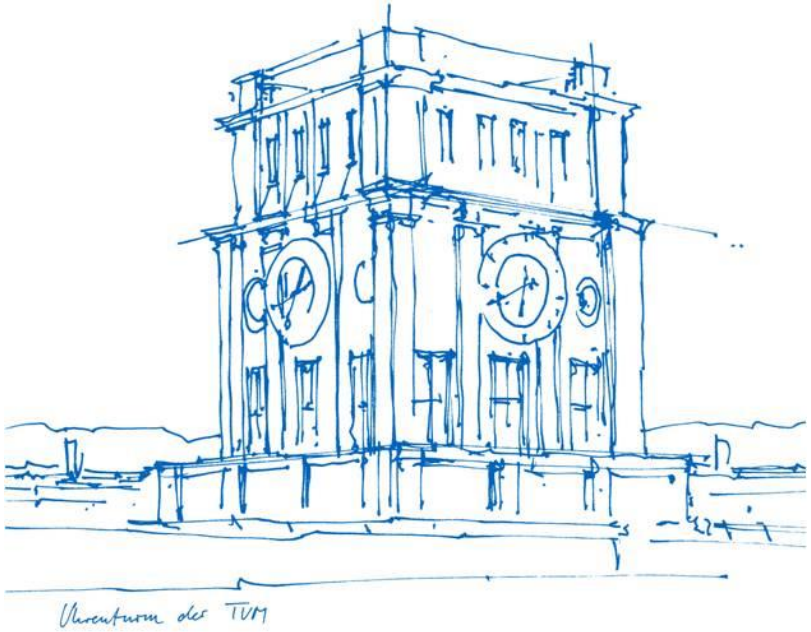
	Author_ID	Author_name	Comment_number	Sentence_number	Domain_Relevance	Sentiment	Entity	Attribute	Sentence	Source_file	Annotator	Aspect
0	Justin-Ma	Justin Ma	521	1	0	NaN	NaN	NaN	Thanks for the thoughtful response.	quora.json	sumit	nan-nan
1	Justin-Ma	Justin Ma	521	2	0	NaN	NaN	NaN	I think we actually have a lot of common groun...	quora.json	sumit	nan-nan
2	Justin-Ma	Justin Ma	521	3	0	NaN	NaN	NaN	All I want to emphasize are my main points: Pr...	quora.json	sumit	nan-nan
3	Justin-Ma	Justin Ma	521	4	9	p	cg	pp	Industrialization is everything about producti...	quora.json	sumit	cg-pp

Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

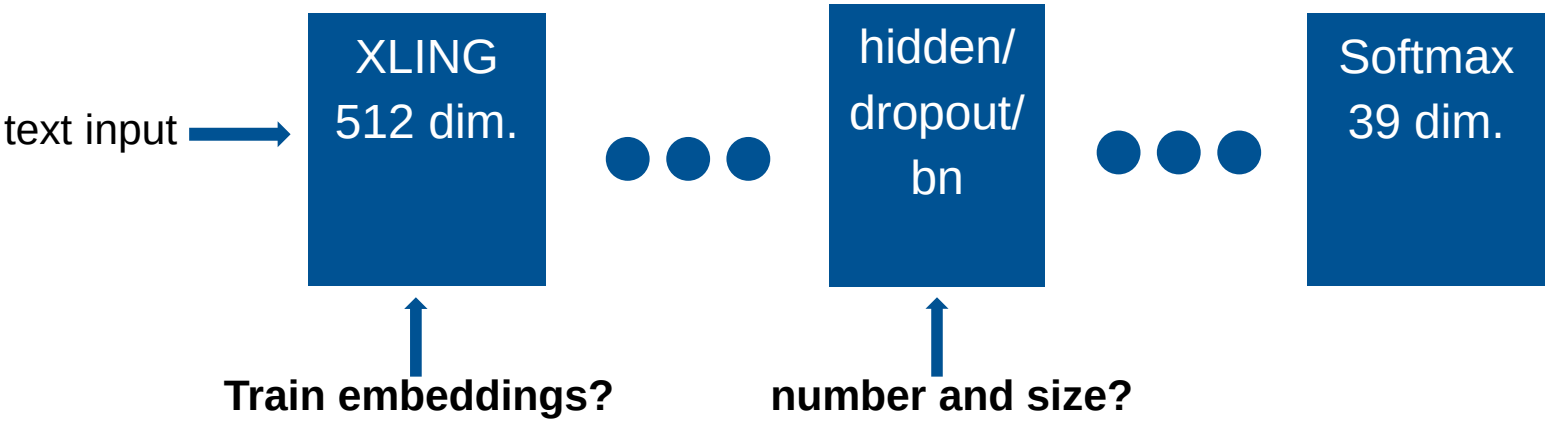
Team 06

Michael Sorg

05.06.2019

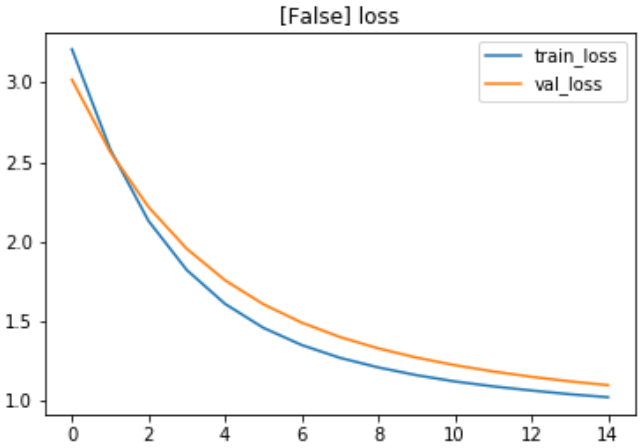


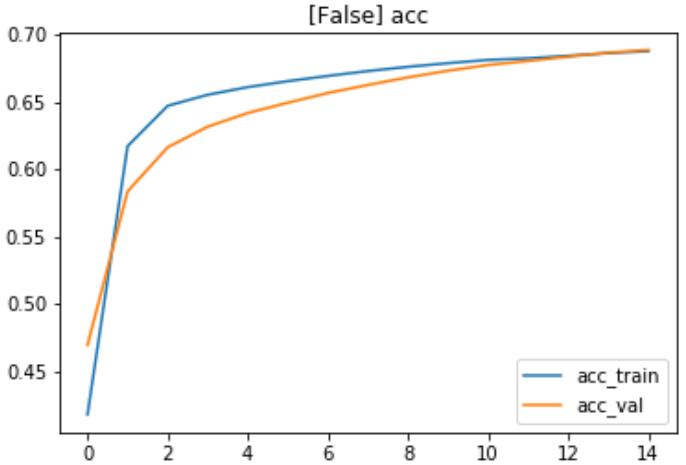
Network search



Baseline experiments

- Amazon Multilingual UK with 100k examples
- One hidden layer (only xling + softmax)

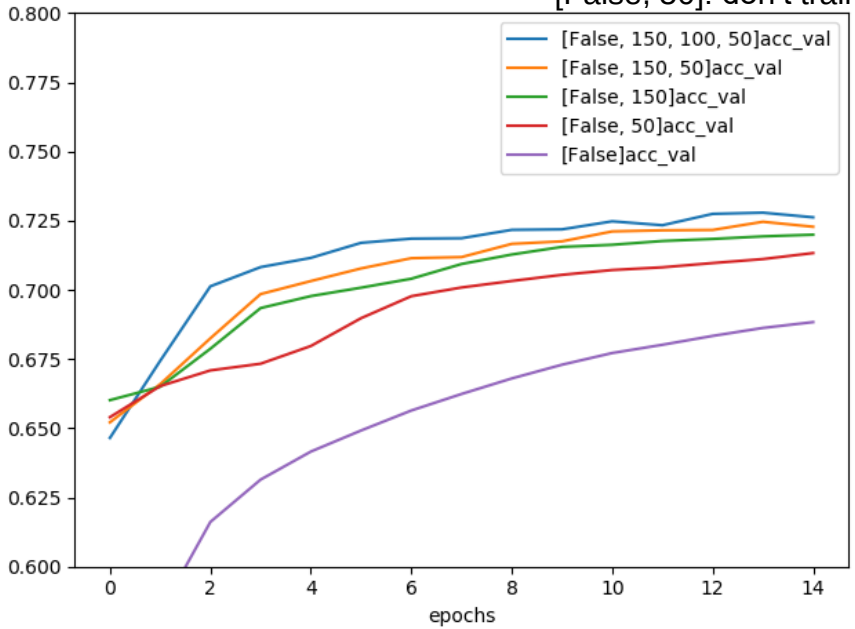




Architecture search

- Amazon Multilingual UK with 100k examples

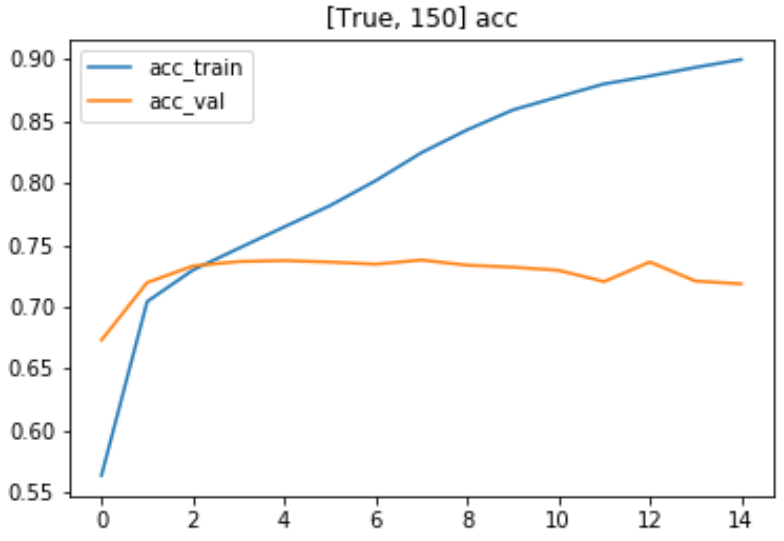
[False, 50]: don‘t train embeddings, fc with 50 units + softmax



Architecture search

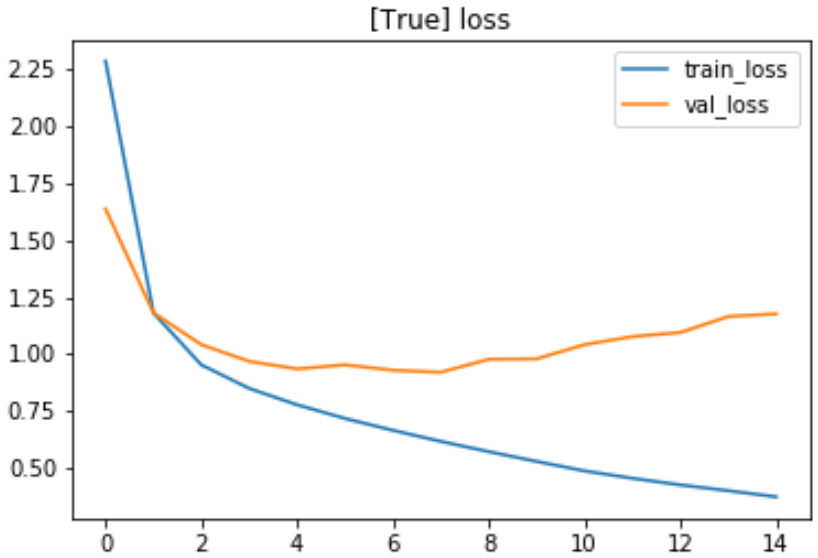
```
self.train_op = tf.train.AdamOptimizer(epsilon=0.1).minimize(self.loss)
```

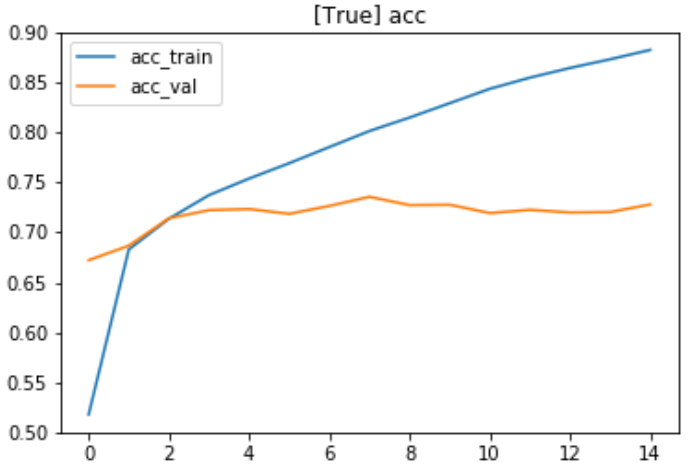
- Train also xling embeddings (100k samples)



Architecture search

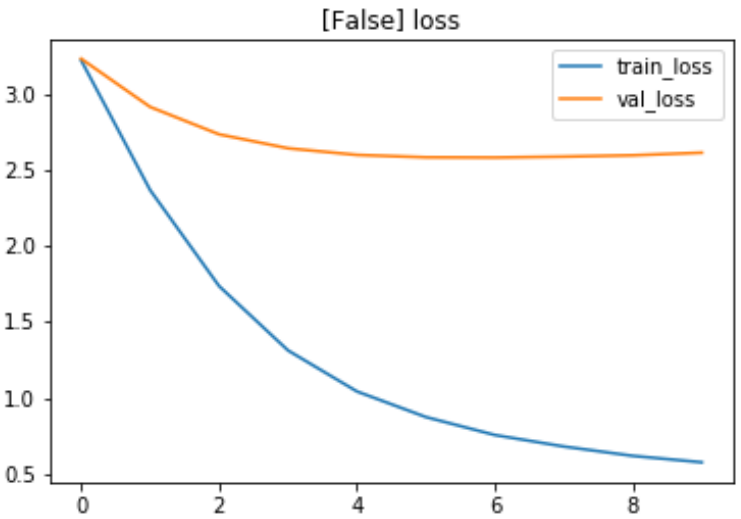
- Train also xling embeddings
- Add dropout to prevent overfitting

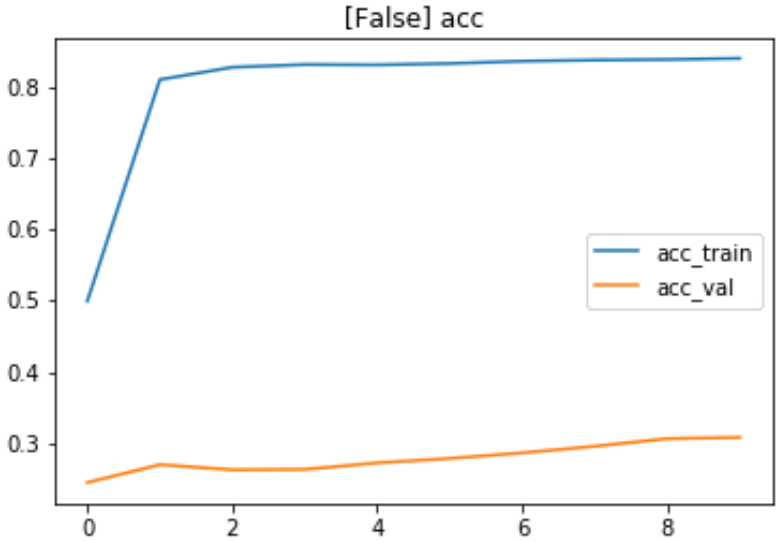




Architecture search

- Amazon Multilingual **US** with 100k examples





```
de.product_category.value_counts()
```

```
Video DVD                  279068
Music                      160588
Books                       63784
Mobile_Apps                 54709
Digital_Video_Download      25124
Digital_Music_Purchase      21554
Toys                        18602
Digital_Ebook_Purchase      12872
PC                          12250
Camera                       5421
Wireless                     4441
Electronics                  4035
Video                        2927
Sports                       2034
Video Games                  1706
Watches                      1575
Shoes                        1517
Home                         1454
Musical Instruments          1094
Baby                          810
Home Improvement              672
Home Entertainment            605
Office Products               412
Personal_Care_Appliances      411
Automotive                    410
Lawn and Garden               397
Luggage                       247
Kitchen                       120
Furniture                      93
Health & Personal Care         37
Software                       19
Pet Products                    2
Grocery                         2
Beauty                          1
Name: product_category, dtype: int64
```

```
[9] us.product_category.value_counts()
```

```
Mobile_Apps                1467128
Digital_Ebook_Purchase     1242173
Video DVD                  1093612
Digital_Video_Download     1051622
Books                       836136
Music                       776810
Digital_Music_Purchase      107461
Toys                         57465
PC                           56817
Video                        46697
Home Entertainment           36298
Wireless                     22637
Camera                       16911
Video Games                  15398
Electronics                  11421
Musical Instruments          10914
Watches                      10537
Tools                         7475
Shoes                         7342
Baby                          5868
Sports                        4123
Home Improvement              3698
Outdoors                      3184
Office Products               2306
Home                          1998
Kitchen                       1834
Lawn and Garden               1204
Health & Personal Care        1066
Automotive                     223
Mobile_Electronics             184
Apparel                        121
Luggage                         78
Beauty                          52
Software                        52
Grocery                         18
Personal_Care_Appliances         9
Furniture                        8
Pet Products                     5
2012-12-22                       1
Name: product_category, dtype: int64
```

```
f = us[0:100000]
f.product_category.value_counts()
```

```
Music              43422
Books              23948
Video              17702
Video DVD          14911
Toys                  10
Office Products        5
Tools                  2
Name: product_category, dtype: int64
```

```
f2 = de[0:100000]
f2.product_category.value_counts()
```

```
Video DVD                  40660
Music                      17961
Mobile_Apps                11905
Digital_Video_Download      6387
Books                       5727
Toys                        3633
Digital_Music_Purchase      3609
Digital_Ebook_Purchase      2579
PC                          2396
Wireless                     876
Camera                       810
Electronics                  779
Video Games                  404
Watches                      324
Shoes                        301
Sports                       267
Musical Instruments          263
Home                         244
Baby                         177
Home Entertainment           133
Home Improvement             124
Automotive                    80
Office Products               78
Lawn and Garden               61
Personal_Care_Appliances      61
Video                         55
Luggage                       51
Furniture                     28
Kitchen                       16
Health & Personal Care         9
Software                       2
Name: product_category, dtype: int64
```

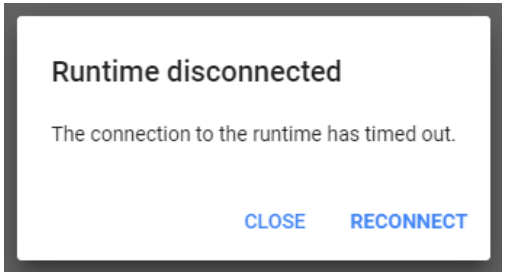
How to find architecture?

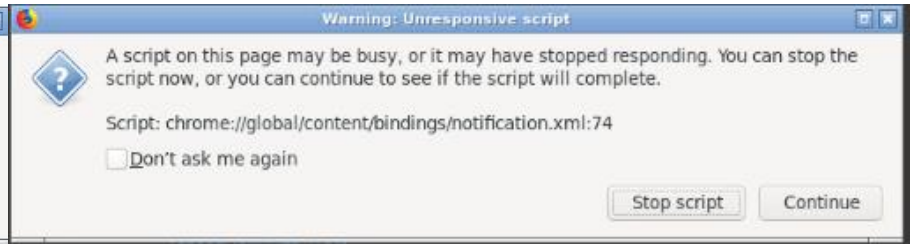
- Use more advanced search techniques (Progressive Neural Architecture Search (PNAS), Efficient Neural Architecture Search (ENAS), Reinforcement learning)

- Tradeoff between time and size of training set ?

Infrastructure Issues

- Training for more than 30 min. fails in most cases



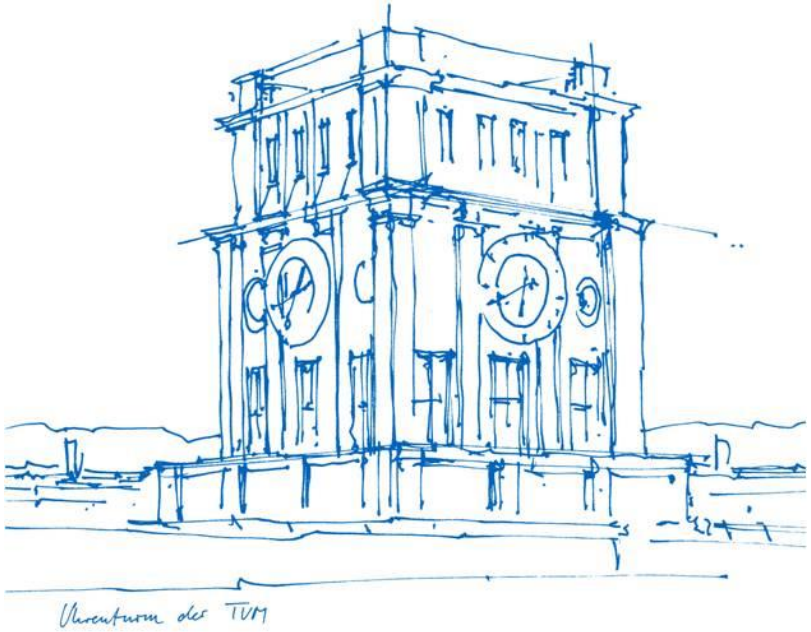


Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

Team 06

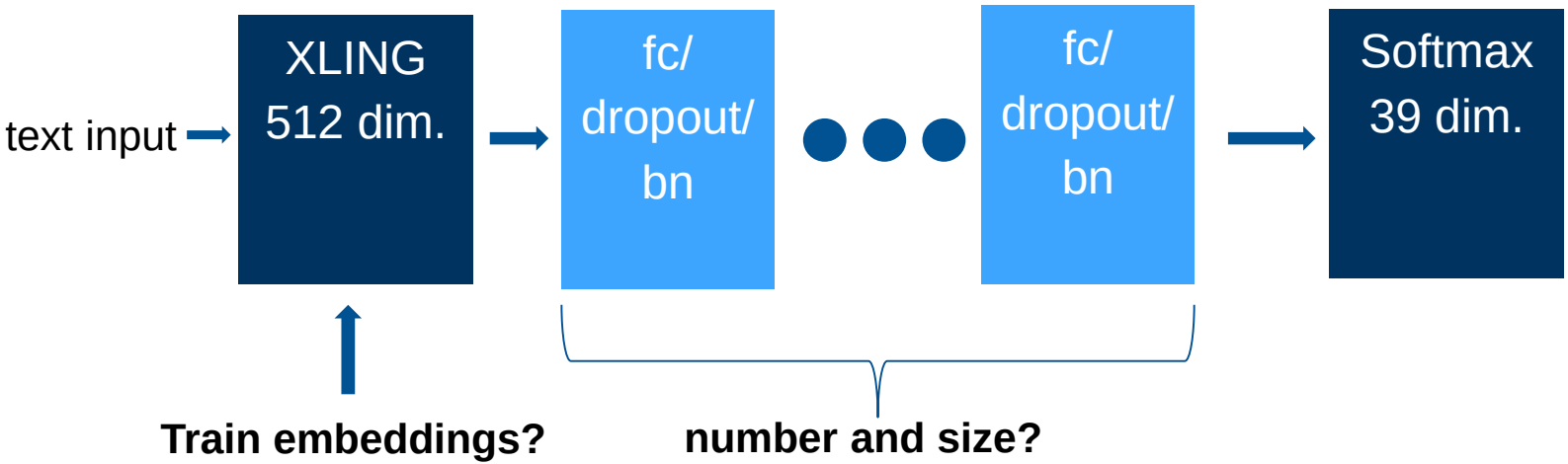
Michael Sorg

19.06.2019



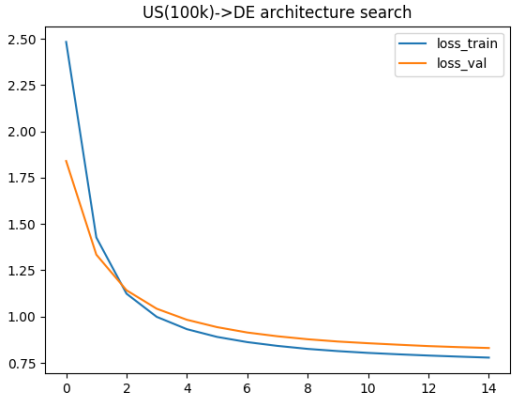
Network search

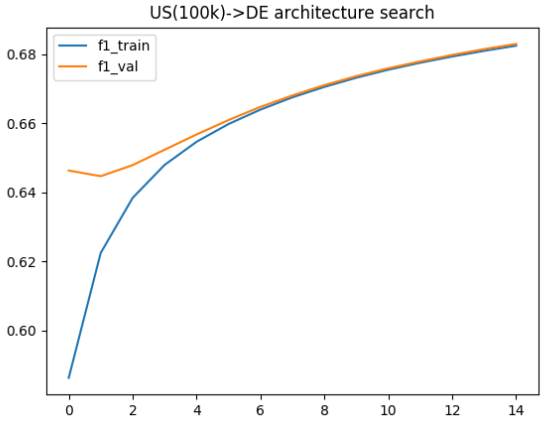
- Task: train on english data only – test on German data

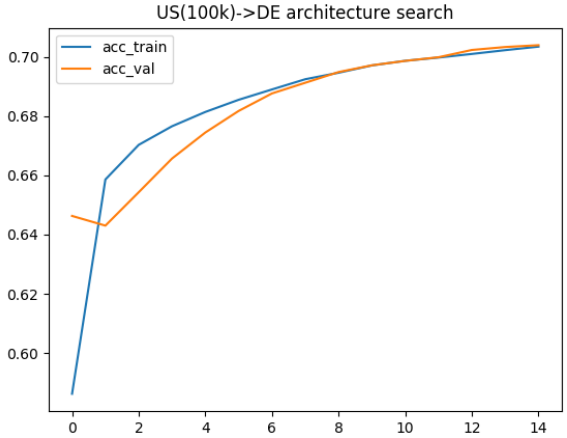


Baseline experiments

- One hidden layer (only xling + softmax)



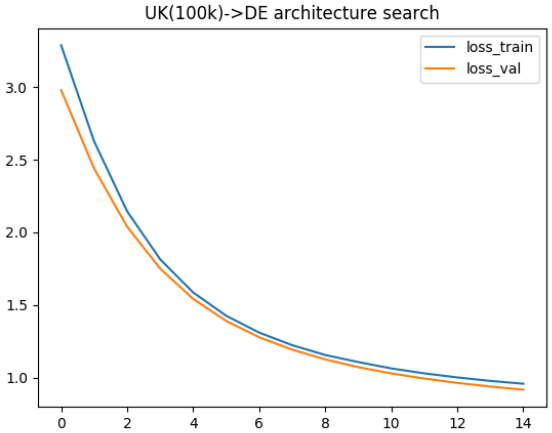


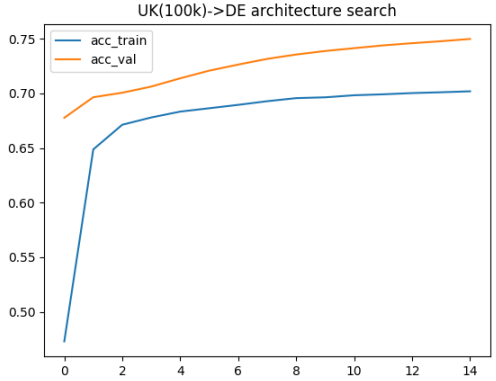


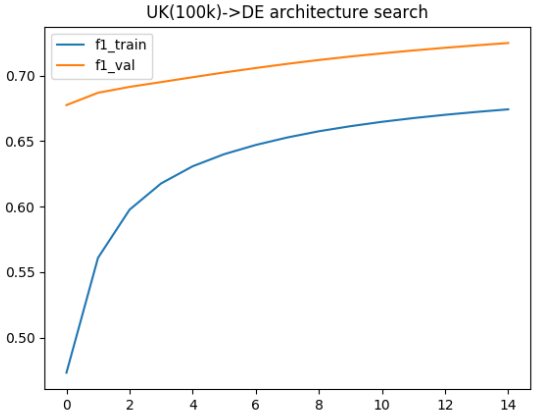
Train data	US (100k)
Test data	DE (100k)

Baseline experiments

- One hidden layer (only xling + softmax)



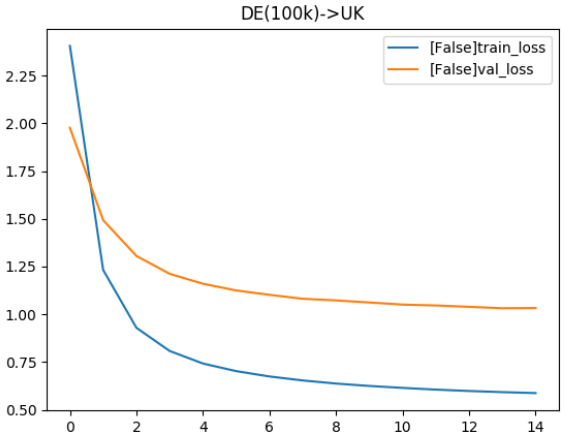


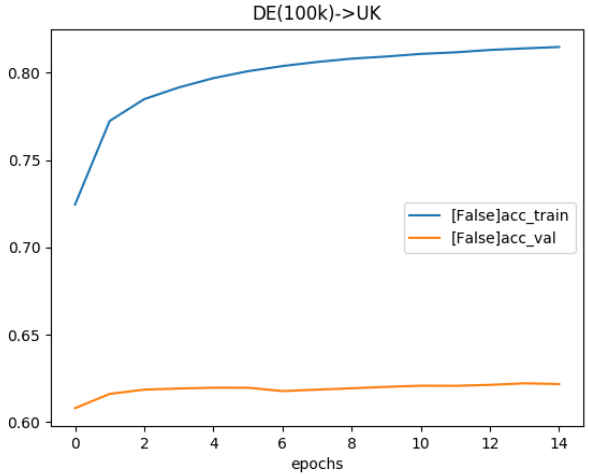


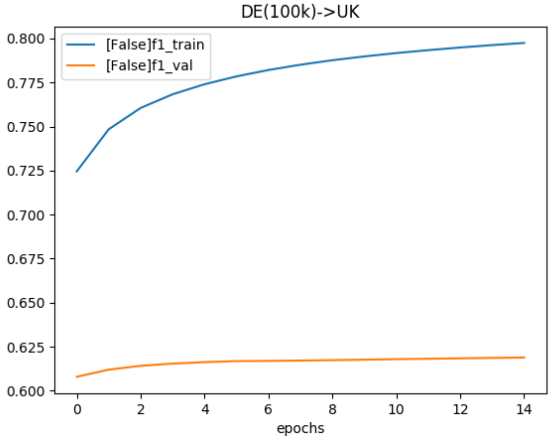
Train data	**UK** (100k)
Test data	DE (100k)

Baseline experiments

- One hidden layer (only xling + softmax)



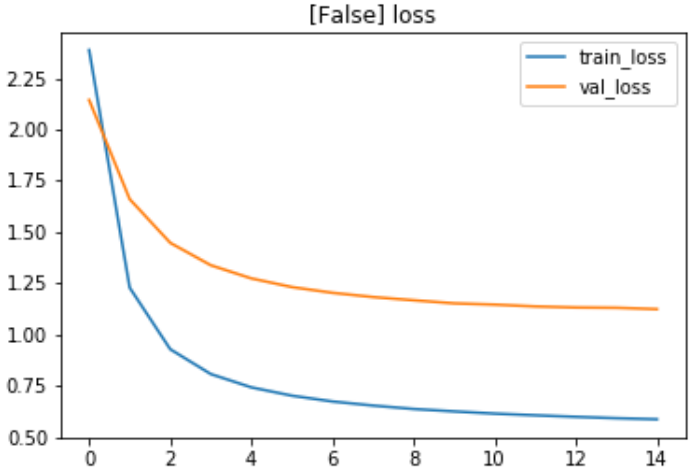


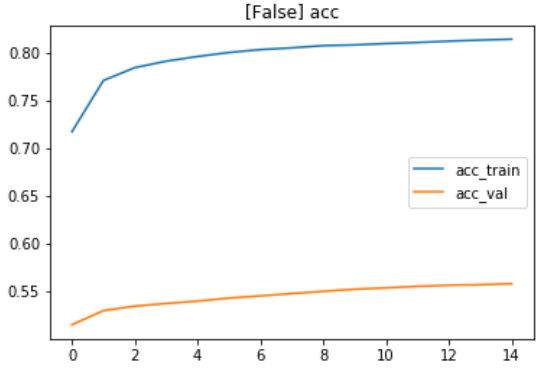


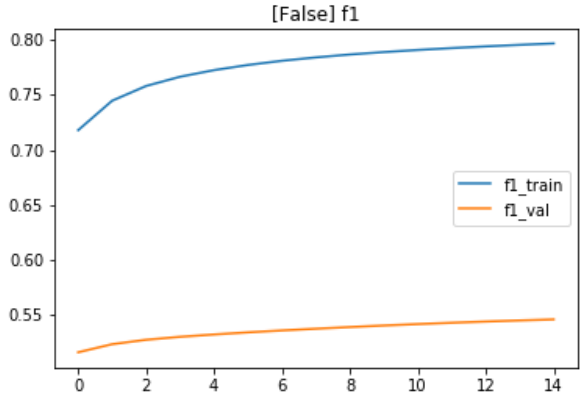
Train data	**DE** (100k)
Test data	UK (100k)

Baseline experiments

- One hidden layer (only xling + softmax)







Train data	**DE** (100k)
Test data	US (100k)

```
de.product_category.value_counts()
```

Video DVD	41048
Music	23890
Books	9353
Mobile_Apps	7998
Digital_Video_Download	3768
Digital_Music_Purchase	3116
Toys	2729
Digital_Ebook_Purchase	1870
PC	1782
Camera	835
Wireless	654
Electronics	566
Video	411
Sports	306
Video Games	247
Watches	238
Home	218
Shoes	202
Musical Instruments	164
Baby	121
Home Improvement	103
Home Entertainment	82
Automotive	70
Lawn and Garden	57
Office Products	52
Personal_Care_Appliances	49
Luggage	27
Kitchen	20
Furniture	15
Health & Personal Care	5
Software	4

Name: product_category, dtype: int64

```
uk.product_category.value_counts()
```

Video DVD	27228
Music	19471
Digital_Ebook_Purchase	16868
Books	15035
Mobile_Apps	12660
Digital_Video_Download	1810
Digital_Music_Purchase	1698
Toys	1507
PC	985
Camera	380
Wireless	353
Electronics	303
Baby	268
Video	256
Video Games	206
Watches	205
Home	179
Musical Instruments	161
Sports	127
Shoes	118
Home Improvement	55
Office Products	53
Automotive	26
Lawn and Garden	18
Health & Personal Care	12
Home Entertainment	5
Software	5
Personal_Care_Appliances	5
Kitchen	1
Pet Products	1
Luggage	1

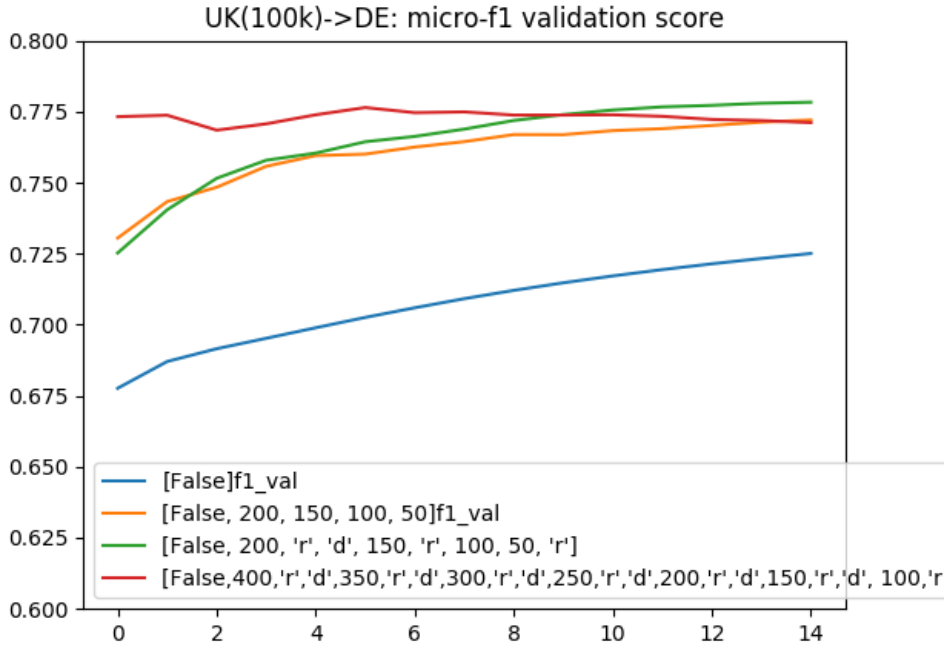
Name: product_category, dtype: int64

```
us.product_category.value_counts()
```

Mobile_Apps	21056
Digital_Ebook_Purchase	18173
Video DVD	15949
Digital_Video_Download	15427
Books	12097
Music	11148
Digital_Music_Purchase	1488
Toys	820
PC	766
Video	666
Home Entertainment	512
Wireless	304
Camera	272
Video Games	226
Musical Instruments	167
Electronics	160
Watches	151
Tools	119
Shoes	111
Baby	100
Sports	63
Outdoors	52
Home Improvement	51
Home	35
Office Products	26
Kitchen	21
Health & Personal Care	15
Lawn and Garden	10
Mobile_Electronics	5
Automotive	4
Luggage	1
Personal_Care_Appliances	1
Grocery	1
Apparel	1
Software	1
Beauty	1

Name: product_category, dtype: int64

Deeper architectures



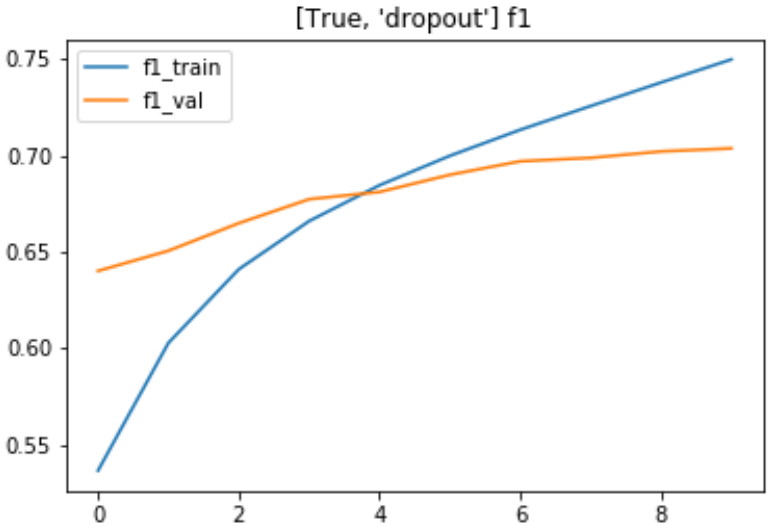
Train data	**UK** (100k)
Test data	DE (100k)

Train Embeddings

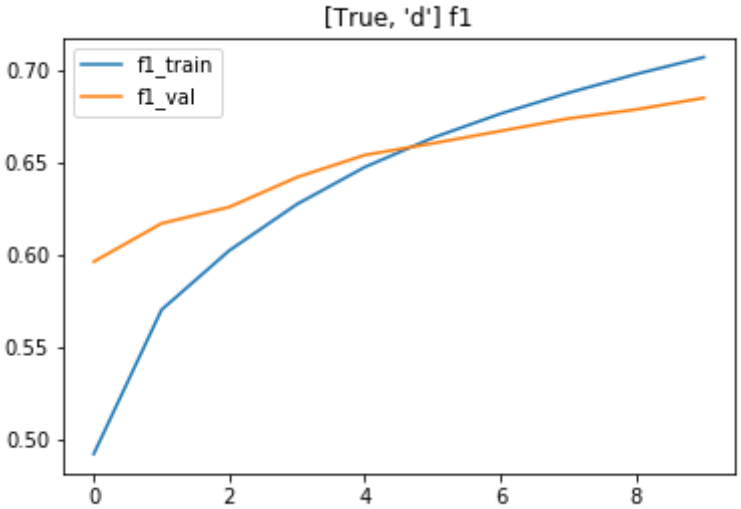
- Quickly overfits
- Even with dropout and reduced learning rate

Train data	**US** (100k)
Test data	DE (100k)

lr: 0.001



lr: 0.0005



Progress

- Colab issues –> switched to google cloud compute engine
- Training 100k examples takes 5 min per epoch (15 epochs ~45 min.)
- Training 1 Million examples takes 30 min per epoch (15 epochs ~7-8 hours)

- Problem: balance between training size and computation time

Roadmap

- Continue with architecture search
- Balanced training

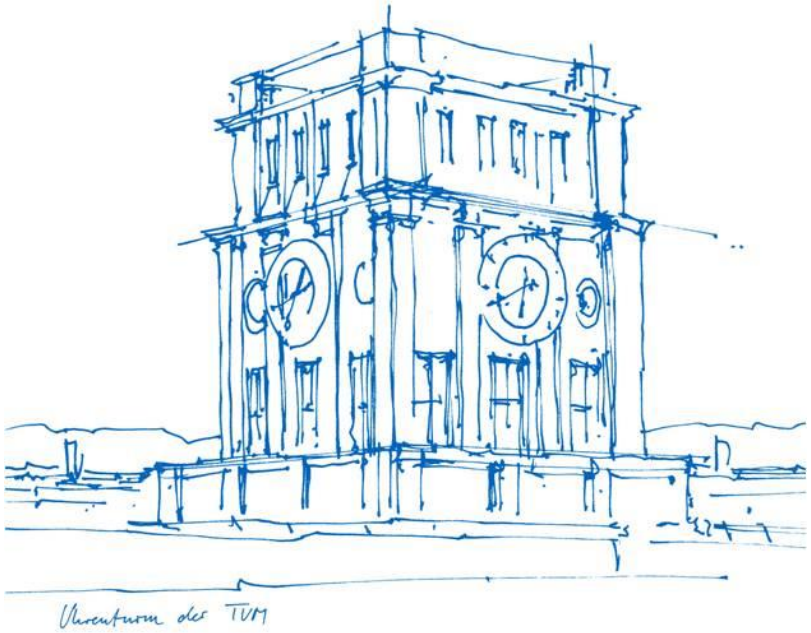
- Run baseline architecture on organic dataset
- Fine-tune on organic dataset

Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

Team 06

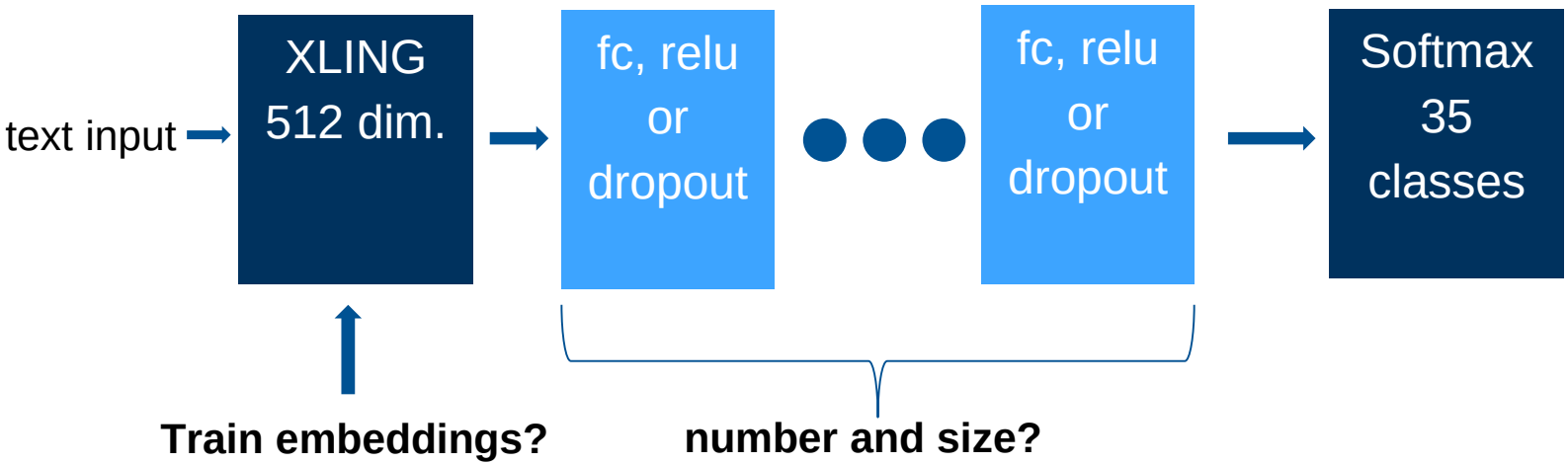
Michael Sorg

03.07.2019



Network search

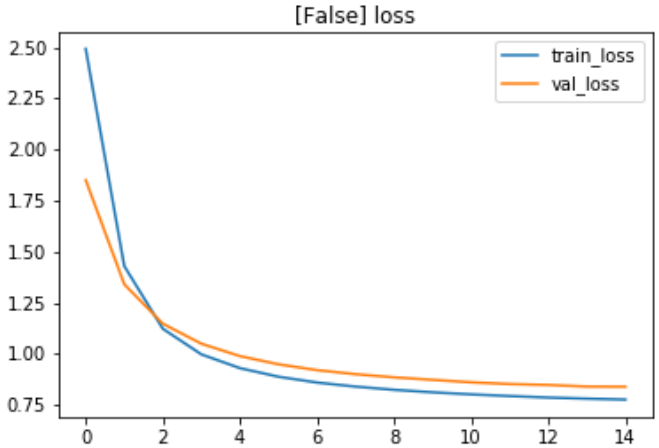
- Task: category prediction on english reviews– test on German

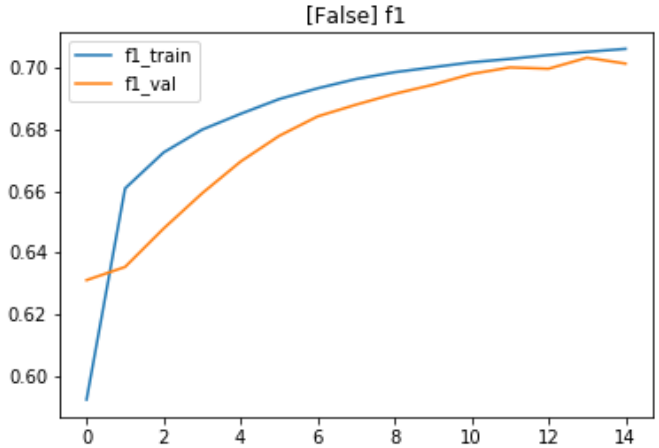


Baseline experiments

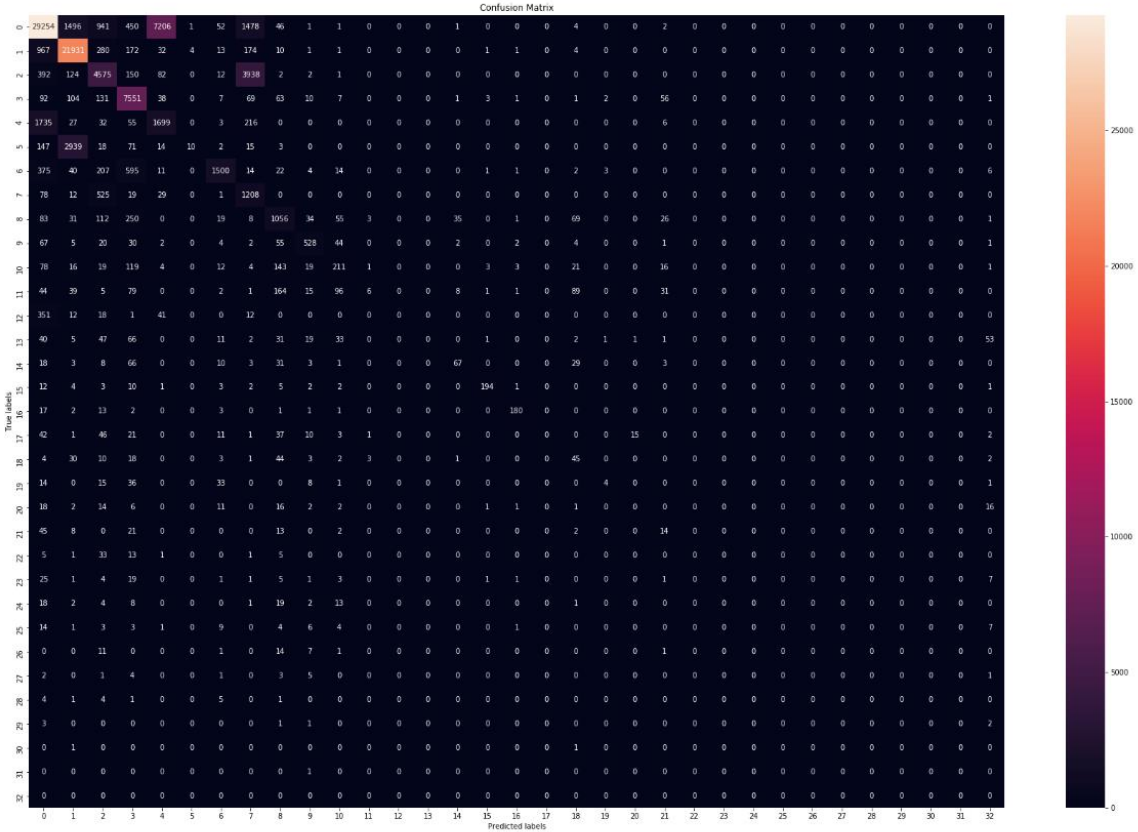
- One hidden layer (only xling + softmax)
- Unbalanced training

Train data	US unbalanced (100k)
Test data	DE (100k)





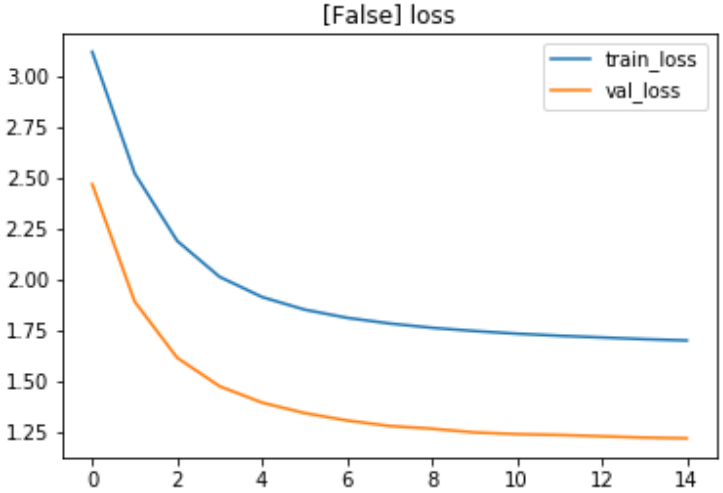
	precision	recall	f1-score	support
0	0.86	0.71	0.78	40933
1	0.82	0.93	0.87	23591
2	0.64	0.49	0.56	9278
3	0.77	0.93	0.84	8137
4	0.19	0.45	0.26	3773
5	0.67	0.00	0.01	3219
6	0.87	0.54	0.66	2795
7	0.17	0.65	0.27	1872
8	0.59	0.59	0.59	1783
9	0.77	0.69	0.73	767
10	0.42	0.31	0.36	670
11	0.43	0.01	0.02	581
12	0.00	0.00	0.00	435
13	0.00	0.00	0.00	313
14	0.58	0.28	0.38	242
15	0.94	0.81	0.87	240
16	0.93	0.82	0.87	220
17	0.00	0.00	0.00	190
18	0.16	0.27	0.20	166
19	0.40	0.04	0.07	112
20	0.00	0.00	0.00	90
21	0.09	0.13	0.11	105
22	0.00	0.00	0.00	59
23	0.00	0.00	0.00	70
24	0.00	0.00	0.00	68
25	0.00	0.00	0.00	53
26	0.00	0.00	0.00	35
27	0.00	0.00	0.00	17
28	0.00	0.00	0.00	16
29	0.00	0.00	0.00	7
30	0.00	0.00	0.00	2
32	0.00	0.00	0.00	1
35	0.00	0.00	0.00	0
micro avg	0.70	0.70	0.70	99840
macro avg	0.31	0.26	0.26	99840
weighted avg	0.75	0.70	0.70	99840

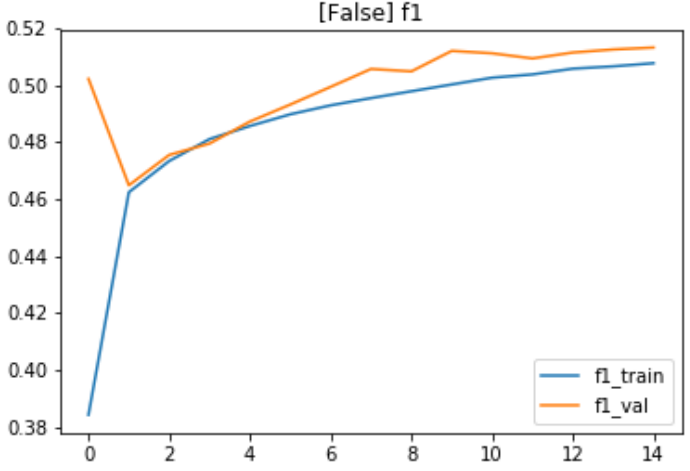


Baseline experiments

- One hidden layer (only xling + softmax)
- **balanced training**

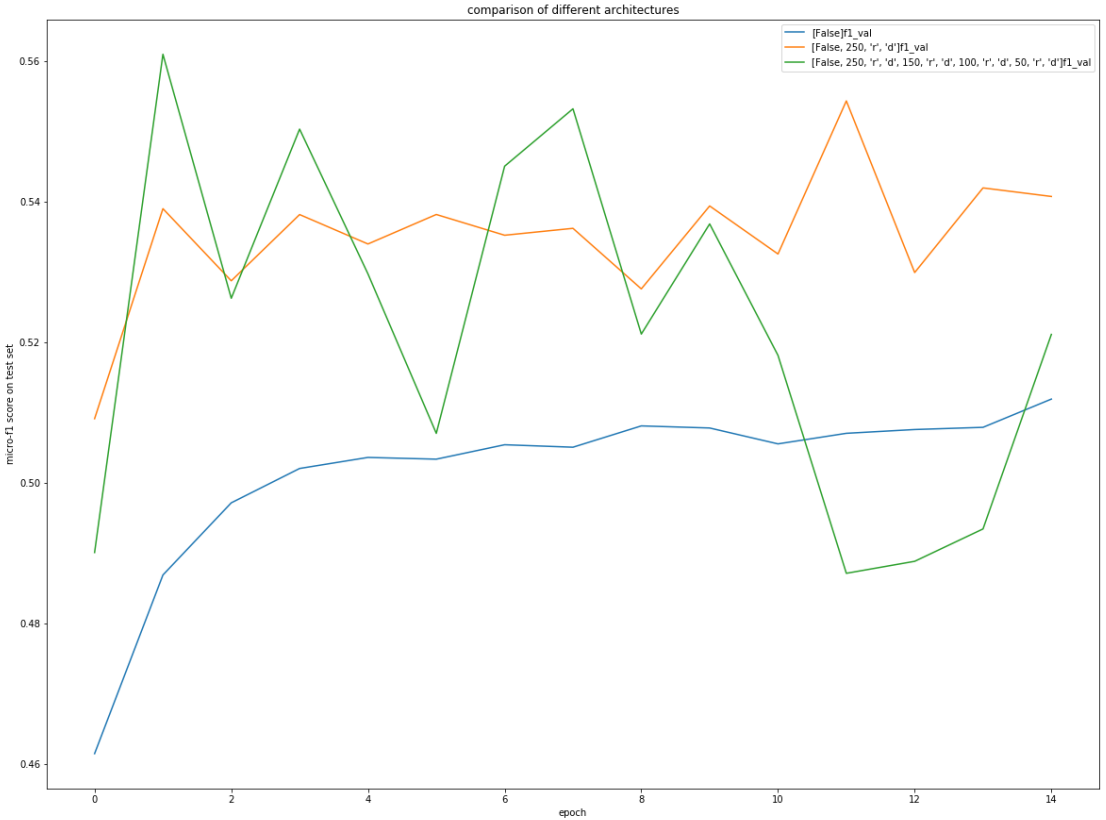
Train data	US **balanced** (100k)
Test data	DE (100k)





	precision	recall	f1-score	support
0	0.90	0.38	0.54	40928
1	0.84	0.84	0.84	23762
2	0.73	0.31	0.44	9446
3	0.92	0.60	0.73	7895
4	0.18	0.32	0.23	3748
5	0.29	0.29	0.29	3158
6	0.75	0.60	0.66	2687
7	0.15	0.71	0.25	1858
8	0.58	0.49	0.53	1843
9	0.58	0.72	0.64	801
10	0.38	0.35	0.37	666
11	0.27	0.35	0.30	553
12	0.02	0.61	0.03	474
13	0.15	0.12	0.13	286
14	0.07	0.47	0.13	250
15	0.60	0.92	0.73	218
16	0.69	0.89	0.78	237
17	0.02	0.01	0.01	246
18	0.11	0.42	0.17	176
19	0.10	0.21	0.13	114
20	0.03	0.07	0.04	88
21	0.03	0.50	0.06	86
22	0.05	0.37	0.09	68
23	0.13	0.36	0.19	64
24	0.07	0.51	0.13	53
25	0.16	0.59	0.25	51
26	0.07	0.63	0.13	38
27	0.03	0.35	0.05	23
28	0.00	0.00	0.00	17
29	0.00	0.00	0.00	5
30	0.00	0.00	0.00	1
31	0.00	0.00	0.00	0
32	0.00	0.00	0.00	0
33	0.00	0.00	0.00	0
accuracy			0.51	99840
macro avg	0.26	0.38	0.26	99840
weighted avg	0.77	0.51	0.58	99840

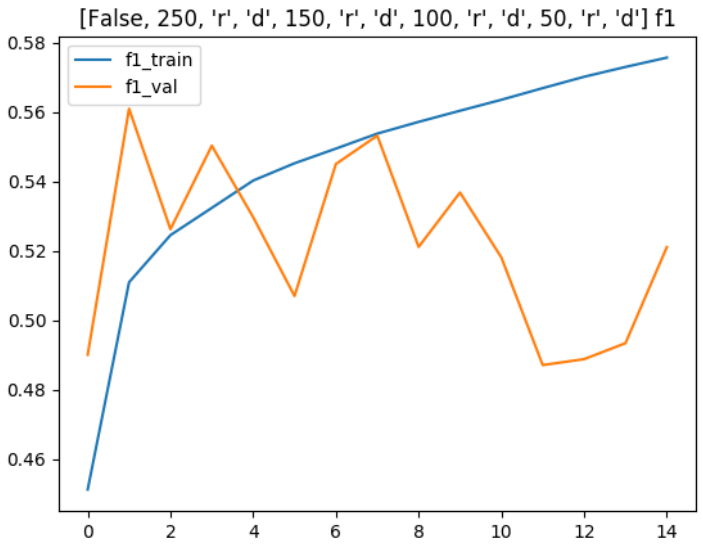
Architecture search



Train data	US balanced (300k)
Test data	DE (300k)

Problems

- Not much learning – f1 score doesn‘t improve much after 2-3 epochs
- zig-zagging f1 score
- Overfitting when using deeper networks (even with dropout & regularization)



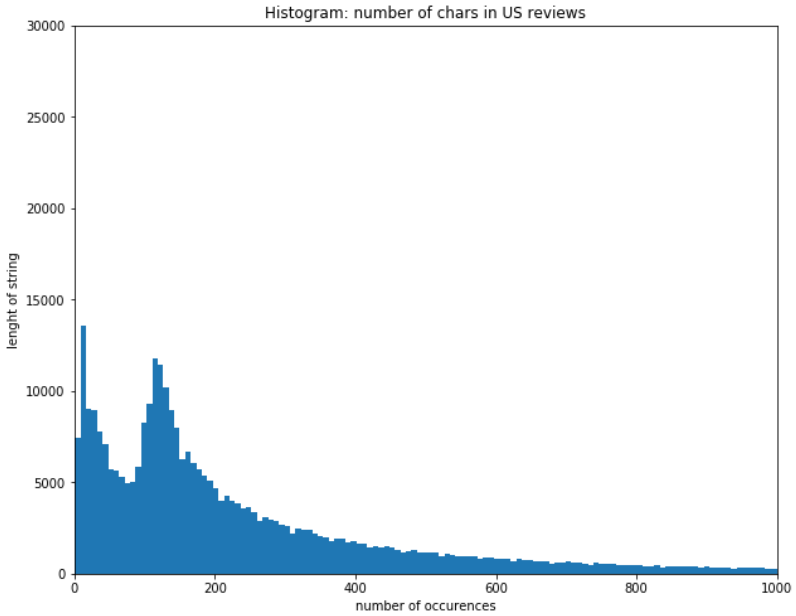
Progress

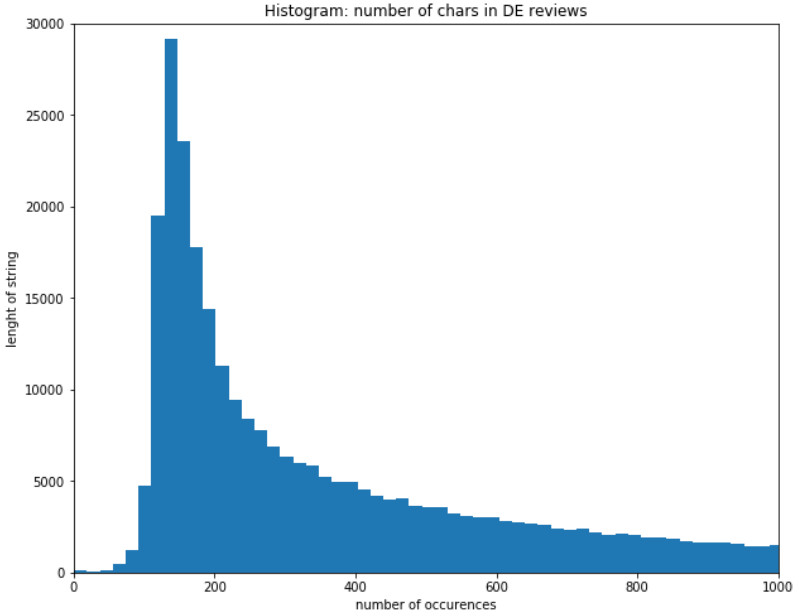
- Tf-idf approach with KNN/SVM (val_f1: 55%, test_f1: 40%)
- size of training set doesn‘t improve performance

Progress

US median: 169.0

- German reviews are longer





Roadmap

- continue with Architecture search

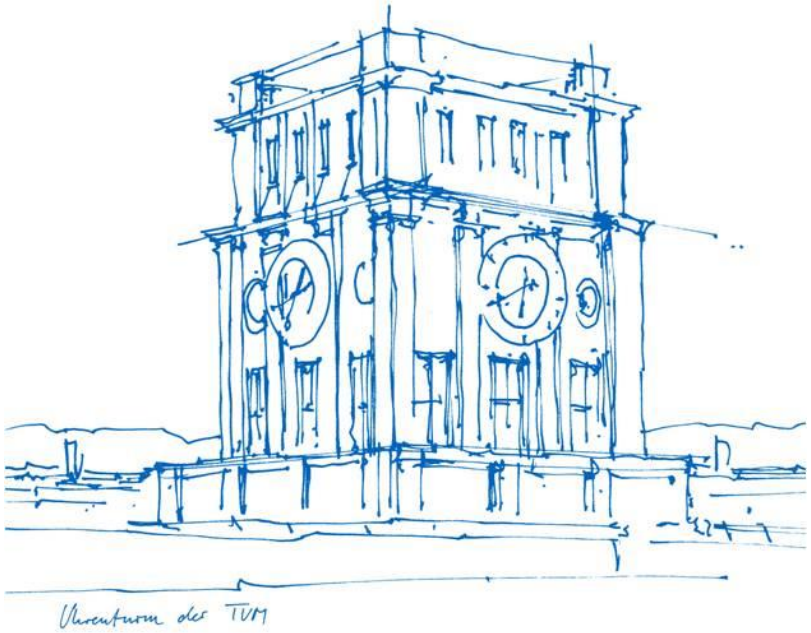
- relevance, entity, and attribute classification on organic dataset
- Fine-tune on organic dataset

Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

Team 06

Michael Sorg

17.07.2019



Outline

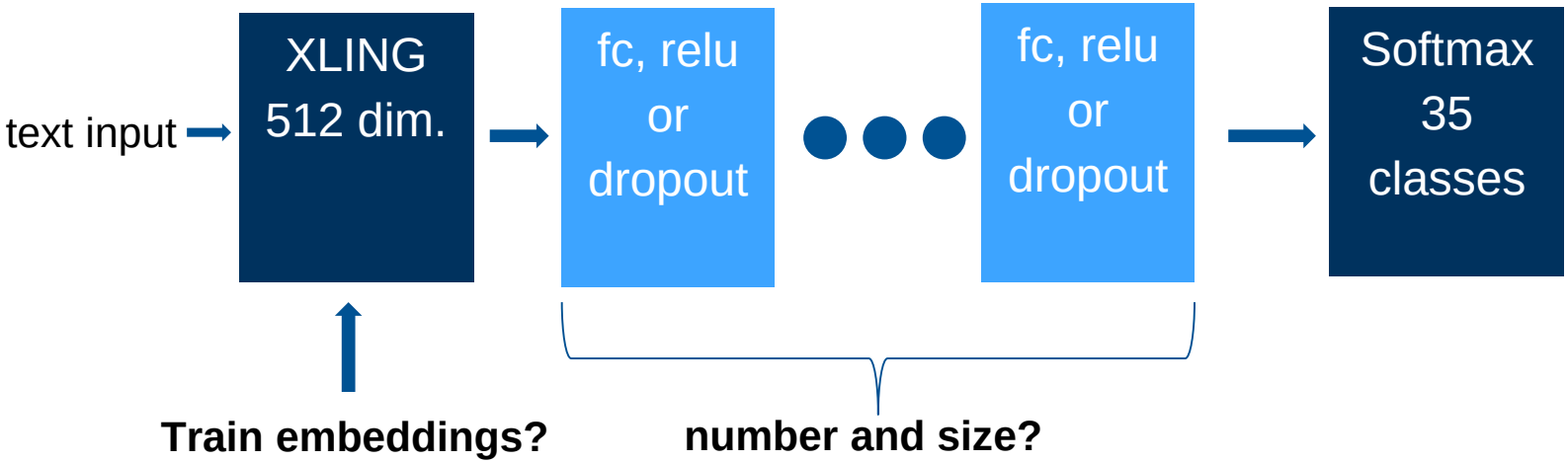
- Task description

- Amazon Reviews
  - Unbalanced training
  - Balanced training
  - Architecture search

- Organic Dataset
  - w/o fine-tuning

Task

- category prediction on amazon reviews based on XLING embeddings
- Evaluate on german reviews without training on german data
- Fine-tune on organic dataset for relevance, attribute and entity classification



Baseline experiment

- no hidden layer (only xling + softmax)
- Unbalanced training

- Some classes are never predicted
- Micro-f1 score on validation set: ~ 70%

	precision	recall	f1-score	support
0	0.86	0.71	0.78	40933
1	0.82	0.93	0.87	23591
2	0.64	0.49	0.56	9278
3	0.77	0.93	0.84	8137
4	0.19	0.45	0.26	3773
5	0.67	0.00	0.01	3219
6	0.87	0.54	0.66	2795
7	0.17	0.65	0.27	1872
8	0.59	0.59	0.59	1783
9	0.77	0.69	0.73	767
10	0.42	0.31	0.36	670
11	0.43	0.01	0.02	581
12	0.00	0.00	0.00	435
13	0.00	0.00	0.00	313
14	0.58	0.28	0.38	242
15	0.94	0.81	0.87	240
16	0.93	0.82	0.87	220
17	0.00	0.00	0.00	190
18	0.16	0.27	0.20	166
19	0.40	0.04	0.07	112
20	0.00	0.00	0.00	90
21	0.09	0.13	0.11	105
22	0.00	0.00	0.00	59
23	0.00	0.00	0.00	70
24	0.00	0.00	0.00	68
25	0.00	0.00	0.00	53
26	0.00	0.00	0.00	35
27	0.00	0.00	0.00	17
28	0.00	0.00	0.00	16
29	0.00	0.00	0.00	7
30	0.00	0.00	0.00	2
32	0.00	0.00	0.00	1
35	0.00	0.00	0.00	0
micro avg	0.70	0.70	0.70	99840
macro avg	0.31	0.26	0.26	99840
weighted avg	0.75	0.70	0.70	99840

Baseline experiment

- no hidden layer (only xling + softmax)
- **balanced training**

- Predictions for some classes are very bad
- Micro-f1 score on validation set: ~ 51%

	precision	recall	f1-score	support
0	0.90	0.38	0.54	40928
1	0.84	0.84	0.84	23762
2	0.73	0.31	0.44	9446
3	0.92	0.60	0.73	7895
4	0.18	0.32	0.23	3748
5	0.29	0.29	0.29	3158
6	0.75	0.60	0.66	2687
7	0.15	0.71	0.25	1858
8	0.58	0.49	0.53	1843
9	0.58	0.72	0.64	801
10	0.38	0.35	0.37	666
11	0.27	0.35	0.30	553
12	0.02	0.61	0.03	474
13	0.15	0.12	0.13	286
14	0.07	0.47	0.13	250
15	0.60	0.92	0.73	218
16	0.69	0.89	0.78	237
17	0.02	0.01	0.01	246
18	0.11	0.42	0.17	176
19	0.10	0.21	0.13	114
20	0.03	0.07	0.04	88
21	0.03	0.50	0.06	86
22	0.05	0.37	0.09	68
23	0.13	0.36	0.19	64
24	0.07	0.51	0.13	53
25	0.16	0.59	0.25	51
26	0.07	0.63	0.13	38
27	0.03	0.35	0.05	23
28	0.00	0.00	0.00	17
29	0.00	0.00	0.00	5
30	0.00	0.00	0.00	1
31	0.00	0.00	0.00	0
32	0.00	0.00	0.00	0
33	0.00	0.00	0.00	0
accuracy			0.51	99840
macro avg	0.26	0.38	0.26	99840
weighted avg	0.77	0.51	0.58	99840

Architecture search

- Training xling embeddings leads to overfitting

- Problems when going deeper:
  - f1 score doesn‘t improve much after 2-3 epochs
  - Overfitting when using deeper networks (more than 2-3 layers)

	micro-f1 on test set
Tf-idf + svm	44 %
Baseline model	51 %
Best model (xling,150,relu,dropout,softmax)	57 %

Results

- Increased data set size doesn‘t lead to better performance

- Including review headline has no effect
- Filtering out English reviews from the German test set has no effect

Organic dataset

- Fine-tuning increases performance

	f1 score	F1 score with fine-tuning
relevance	74 %	77 %
entity	51 %	57 %
attribute	44 %	50 %