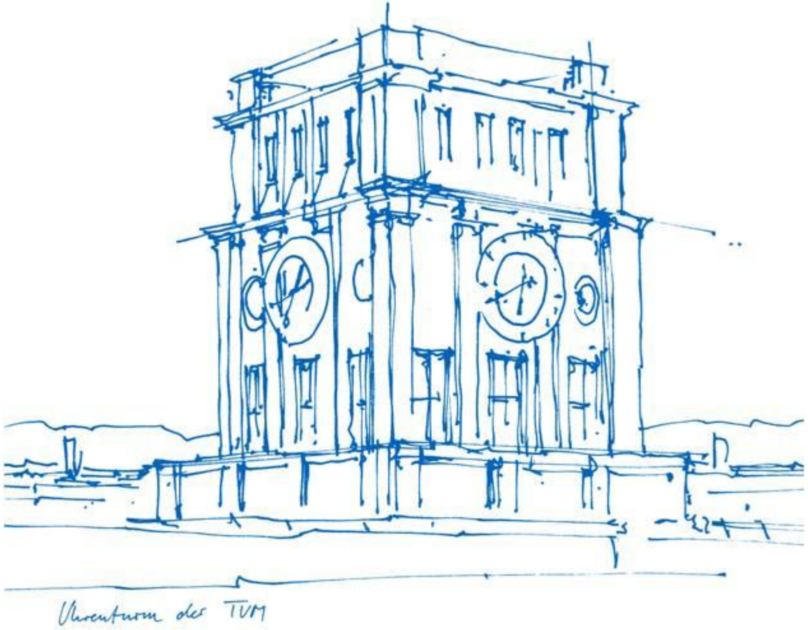
Using the Explainable AI SS3 Classification Model

Muneeb Vaiyani, Florian Angermeir

Technical University Munich

Department of Informatics

Munich, 20. July 2020



SS3

- Simple Text Classifier
- Published in 2019 [1]
- Key Aspects
  - Incremental classification of sequential data
  - Support for early classification
  - Explainability

[1] Burdisso et al. "A text classification framework for simple and effective early depression detection over social media streams." *Expert Systems with Applications* 133 (2019): 182-197.

SS3 - Training

SS3 - Training

Training Set

Text	Class
This sample is a unique sample	Class 1
This sample is great	Class 1
Works like a charm	Class 2
Works with a sample	Class 2

Dict Class 1

Term	Freq.
sample	3
is	2
this	2
a	1
great	1
unique	1

Dict Class 2

Term	Freq.
works	2
a	2
sample	1
like	1
charm	1
with	1

SS3 - Classification

SS3 - Classification

1. Split sample up

“This sample is a great example. This algorithm works like a charm.”

“This sample is a great example.”

“This algorithm works like a charm.”

“This” “sample” “is” “a” “great” “example”

“This” “algorithm” “works” “like” “a” “charm”

SS3 - Classification

2. Calculate confidence vectors for n-grams (here 1-grams)

confidence vector: (class 1, class 2)

confidence value for class 1

confidence value for class 2

confidence value: how certain is SS3 that word exclusively belongs to class

"This"	*"sample"*	*"is"*	*"a"*	*"great"*	*"example"*
(0.83,0)	(1.0,0)	(0.83,0)	(0,1.0)	(0,0)	(0,0)

"This"	*"algorithm"*	*"works"*	*"like"*	*"a"*	*"charm"*
(0.83,0)	(0,0)	(0,0)	(0,0)	(0,1.0)	(0,0)

SS3 - Classification

2. Calculate confidence vectors for n-grams (here 1-grams)

$$confidence\ value(class, word) = local\ value(class, word) \cdot significance(class, word) \cdot sanction(class, word)$$

local value: proportionally to term frequency in class
significance: importance of word for a class
sanction: proportionally decreases in relation to # of class a word is significant to

SS3 - Classification

3. Summarize confidence vectors

“This” “sample” “is” “a” “great” “example”

(0.83,0) (1.0,0) (0.83,0) (0,1.0) (0,0) (0,0)

⊕

(2.66, 1.0)

“This” “algorithm” “works” “like” “a” “charm”

(0.83, 0) (0,0) (0,0) (0,0) (0,1.0) (0,0)

⊕

(0.83, 1.0)

⊕

(3.49, 2.0)

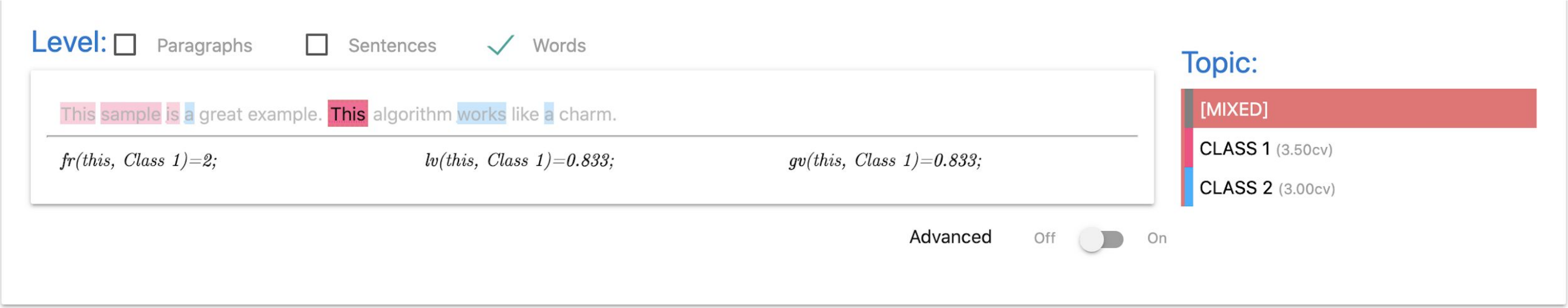
⇨ **Class 1**

SS3 - Explainability

SS3 - Explainability

Only four hyperparameters ($confidence\ value(class, word) = local\ value(class, word) \cdot significance(class, word) \cdot sanction(class, word)$)

- α in confidence value
- σ in local value
- ƛ in significance
- ρ in sanction



Phase 1

Phase 1

Goal:

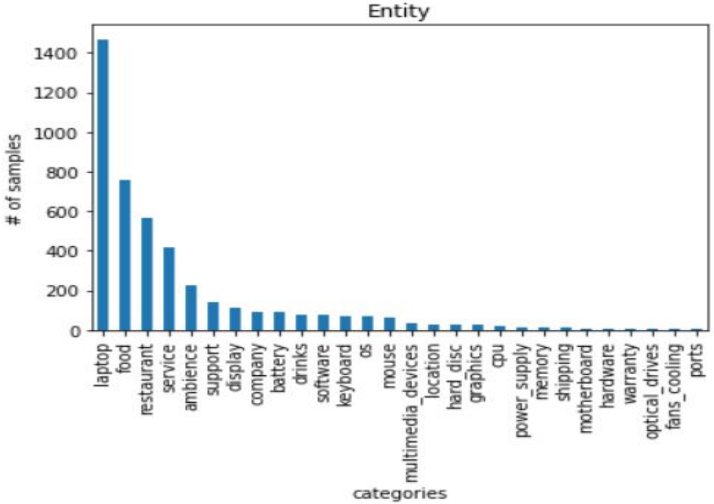
- Find out whether SS3 works for entity/attribute classification
    - Use SemEval 2016 Task 5 Dataset [1]
- If works
    - Analyze reasoning of classifier
    - Apply to Organic Food Dataset

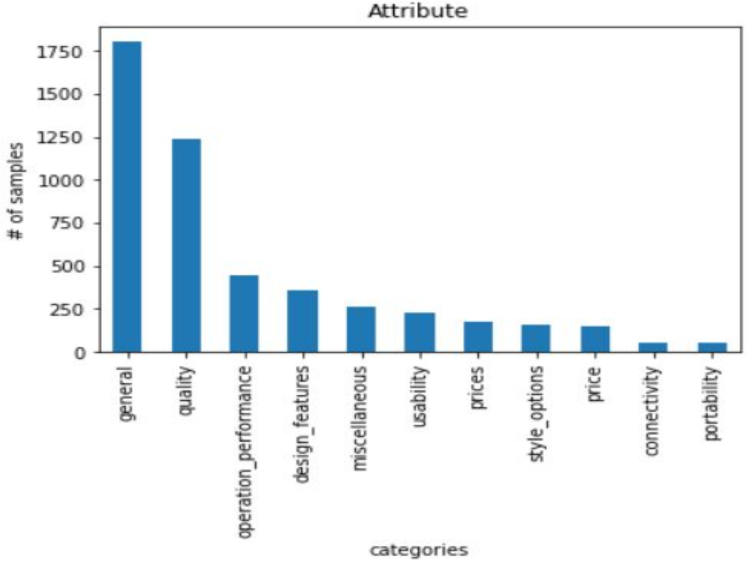
[1] http://alt.qcri.org/semeval2016/task5/

Phase 1

Dataset:

- SemEval 2016 Task 5 Laptop & Restaurant Dataset combined
- Multi-labeled





Phase 1

Default 3-gram classification

- Initially SS3 did not support multilabel classification [1,2]
- Optimized the hyperparameters for macro average f1 score

Entity/Attribute	Macro F1	Micro F1
Entity	52 %	58 %
Attribute	53 %	62 %

[1] https://github.com/sergioburdisso/pyss3/issues/6
[2] https://github.com/sergioburdisso/pyss3/issues/5

Phase 1

Approach (Attribute)	Impr. Macro F1	Impr. Micro F1
Average Operator	- 2 %	- 3 %
Maximum Operator	4 %	- 9 %
Reduce Classes	9 %	1 %
Word Embeddings	2 %	0 %
Bigger Sample Size	- 16 %	- 3 %

Phase 2

Phase 2

Goal:

- Improve classifier’s architecture
- Find dataset characteristics that might have influence on SS3 performance

Phase 2

Improve Classifier

	Impr. Macro F1	Impr. Micro F1
One vs. Rest Classification [1,2]	<= 4 %	<= 16 %
Class Specific Hyperparameters [3,4]	<= 5 %	<= 8 %

Datasets

[1] http://www.cs.cmu.edu/~ark/personas/

[2] http://alt.qcri.org/semeval2016/task5/

[3] http://alt.qcri.org/semeval2017/task4/

[4] https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews/data

Phase 2

Dataset Characteristics

	Avg #Word	Avg #Sentence	Avg #Paragraph	#Classes	Imbalance	Avg Imp. Word Ratio	Word Overlap Ratio	Important Word Overlap Ratio	Macro F1 Score
Spam	18.892529	2.007853	1.000449	2.0	6.440735	0.442924	0.148323	0.059447	0.921793
MovieReview	267.356007	12.552728	5.114754	2.0	1.575299	0.439056	0.495548	0.433929	0.861986
IMDB	270.112381	12.513696	5.123446	2.0	1.015178	0.443151	0.562934	0.315599	0.830358
TopicClassification	28.458472	7.945024	1.000000	8.0	1.868425	0.537998	0.302252	0.302000	0.694964
Apple	20.000767	1.562883	1.134969	3.0	5.350000	0.149534	0.197989	0.197989	0.673001
Beer	6.638787	0.999865	1.000000	7.0	2.199052	0.791577	0.458781	0.458781	0.623486
SemEval	23.801818	1.746259	1.000000	3.0	3.194036	0.136304	0.450804	0.449308	0.517649
Alexa5	29.020238	2.300000	1.000000	5.0	23.740260	0.213439	0.397746	0.397746	0.408180
Alexa2	29.118651	2.317063	1.000000	2.0	6.522388	0.384660	0.347804	0.248004	0.408180
Clothing5	66.848361	4.645252	1.194965	5.0	15.449339	0.171963	0.592775	0.592775	0.311794
Clothing2	66.910794	4.644241	1.201139	2.0	4.660741	0.432836	0.463526	0.421737	0.311794
City	16.481412	1.079985	1.000000	24.0	179.250000	0.461586	0.453799	0.453799	0.123188

[1] https://gitlab.lrz.de/nlp-lab-course-ss2020/opinion-mining/opinion-lab-group-1.5/-/wikis/03.-Phase-2

Phase 3

Phase 3

Goal:

- Analyze reasoning of the entity/attribute classification of beer and city review

Datasets:

1. **Beer review [1]**
2. **City review [1]**

[1] https://github.com/ruidan/Unsupervised-Aspect-Extraction

Phase 3

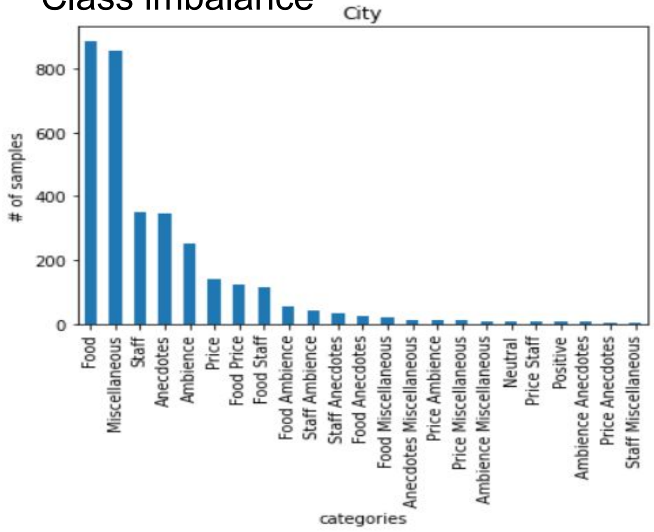
Why is beer performing better than city?

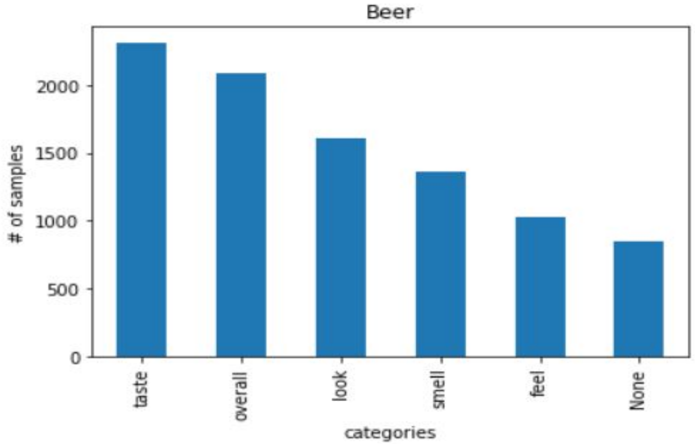
	Macro F1 Score	Micro F1 Score
Beer Review	76 %	76%
City Review	12 %	36%

Phase 3

- Few classes in beer corpus compared to city one.

- Class imbalance





Phase 3

- More samples per class ⟶ bigger vocabulary per class

Category	Index	Length	Vocab. Size
None	0	3902	1463
feel	1	4834	1007
look	2	9586	1490
overall	3	10521	2466
smell	4	7207	1472
taste	5	12983	2347

Category	Index	Length	Vocab. Size
Ambience	0	3216	948
Ambience Anecdotes	1	134	95
Ambience Miscellaneous	2	164	119
Anecdotes	3	4328	1027
Anecdotes Miscellaneous	4	259	140
Food	5	10614	2080
Food Ambience	6	865	387
Food Anecdotes	7	332	208
Food Miscellaneous	8	344	202
Food Price	9	1701	546
Food Staff	10	1550	539
Miscellaneous	11	7325	1489
Neutral	12	99	78
Positive	13	106	76
Price	14	1343	444
Price Ambience	15	159	108
Price Anecdotes	16	63	49
Price Miscellaneous	17	210	132
Price Staff	18	84	55
Staff	19	4698	1168
Staff Ambience	20	544	238
Staff Anecdotes	21	508	256
Staff Miscellaneous	22	90	73

Phase 3

- Vocabulary learned by classifier has low overlap

Frequency in dataset

Importance to classifier

- Beer (taste): Frequent words also has high global value
- City (staff): Frequent words do not have high global value

Phase 3

Deep diving into the few false negatives of city review (True Class: Staff, False Positive: Food)

1. In **an** effort to increase turnover , the restaurant **offers no** **desserts** beyond the **complimentary** espresso cup **filled** with **chocolate** mousse .

2. **Only** drawback - **they** wo n't **toast your bagel** , and **they do n't make** **eggs** for the **bagel** .

3. The **service** is not **consistently** **excellent** -- **just** **decent** .

word: **(Staff cv > Food cv)**
word: **(Staff cv < Food cv)**
word: **(word does not exist for staff but exist for food)**

Conclusion

Can SS3 be used for entity/attribute classification?

- Partially, depends on dataset (work better for less classes)
- SS3 not powerful enough to capture the necessary context since solely depends on term frequency
- Multilabel Classification Tasks too hard (inherent overlap)
- Better for dataset with more distinctive words for each class.
- Architectural changes (e.g. policy for selecting # multi label (k_means) might improve classification)

Analysis of reasoning

- SS3 directly supports sample wise explainability
- Scaling of explainability to whole dataset/task needs a lot of additional effort

Backup

SS3

Local Value

Dict Class 1

Term	Freq.
sample	3
is	2
this	2
a	1
great	1
unique	1

Dict Class 2

Term	Freq.
works	2
a	2
sample	1
like	1
charm	1
with	1

$$lv_\sigma(w,c) = \left(\frac{tf_{w,c}}{max\{tf_c\}}\right)^\sigma$$

Local Value Class 1

Term	LV
sample	1.0
is	0.83
this	0.83
a	0.60
great	0.60
unique	0.60

Local Value Class 2

Term	LV
works	1.0
a	1.0
sample	0.73
like	0.73
charm	0.73
with	0.73

Significance

Local Value Class 1

Term	LV
sample	1.0
is	0.83
this	0.83
a	0.60
great	0.60
unique	0.60

Local Value Class 2

Term	LV
works	1.0
a	1.0
sample	0.73
like	0.73
charm	0.73
with	0.73

$$sg_\lambda(w,c) = \frac{1}{2} tanh\Big(4\frac{(lv(w,c) - \widetilde{LV}_w)}{\lambda \cdot MAD_w} - 2\Big) + \frac{1}{2}$$

$$\tilde{LV}_w = median(\mid LV_w \mid)$$

$$MAD_w = median(\mid lv(w,c_i) - \tilde{LV}_w \mid)$$

$$LV_w = \{lv(w,c_i) | c_i \in C\}$$

Significance Class 1

Term	SG
sample	1.0
is	1.0
this	1.0
a	0
great	0
unique	0

Significance Class 2

Term	SG
works	1.0
a	1.0
sample	0
like	0
charm	0
with	0

Sanction

Significance Class 1

Term	SG
sample	1.0
is	1.0
this	1.0
a	0
great	0
unique	0

Significance Class 2

Term	SG
works	1.0
a	1.0
sample	0
like	0
charm	0
with	0

$$sn_\rho(w,c) = \left(-\frac{\hat{C}_{wc}}{|C|-1}+1\right)^\rho$$

$$\hat{C}_{wc} = \sum_{c_i \in C-\{c\}} sg_\lambda(w,c_i)$$

Sanction Class 1

Term	SN
sample	1.0
is	1.0
this	1.0
a	0
great	1.0
unique	1.0

Sanction Class 2

Term	SN
works	1.0
a	1.0
sample	0
like	1.0
charm	1.0
with	1.0

Confidence Value

Local Value Class 1

Term	LV
sample	1.0
is	0.83
this	0.83
a	0.60
great	0.60
unique	0.60

Local Value Class 2

Term	LV
works	1.0
a	1.0
sample	0.73
like	0.73
charm	0.73
with	0.73

Significance Class 1

Term	SG
sample	1.0
is	1.0
this	1.0
a	0
great	0
unique	0

Significance Class 2

Term	SG
works	1.0
a	1.0
sample	0
like	0
charm	0
with	0

Sanction Class 1

Term	SN
sample	1.0
is	1.0
this	1.0
a	0
great	1.0
unique	1.0

Sanction Class 2

Term	SN
works	1.0
a	1.0
sample	0
like	1.0
charm	1.0
with	1.0

$$gv(w,c) = lv_\sigma(w,c) \cdot sg_\lambda(w,c) \cdot sn_\rho(w,c)$$

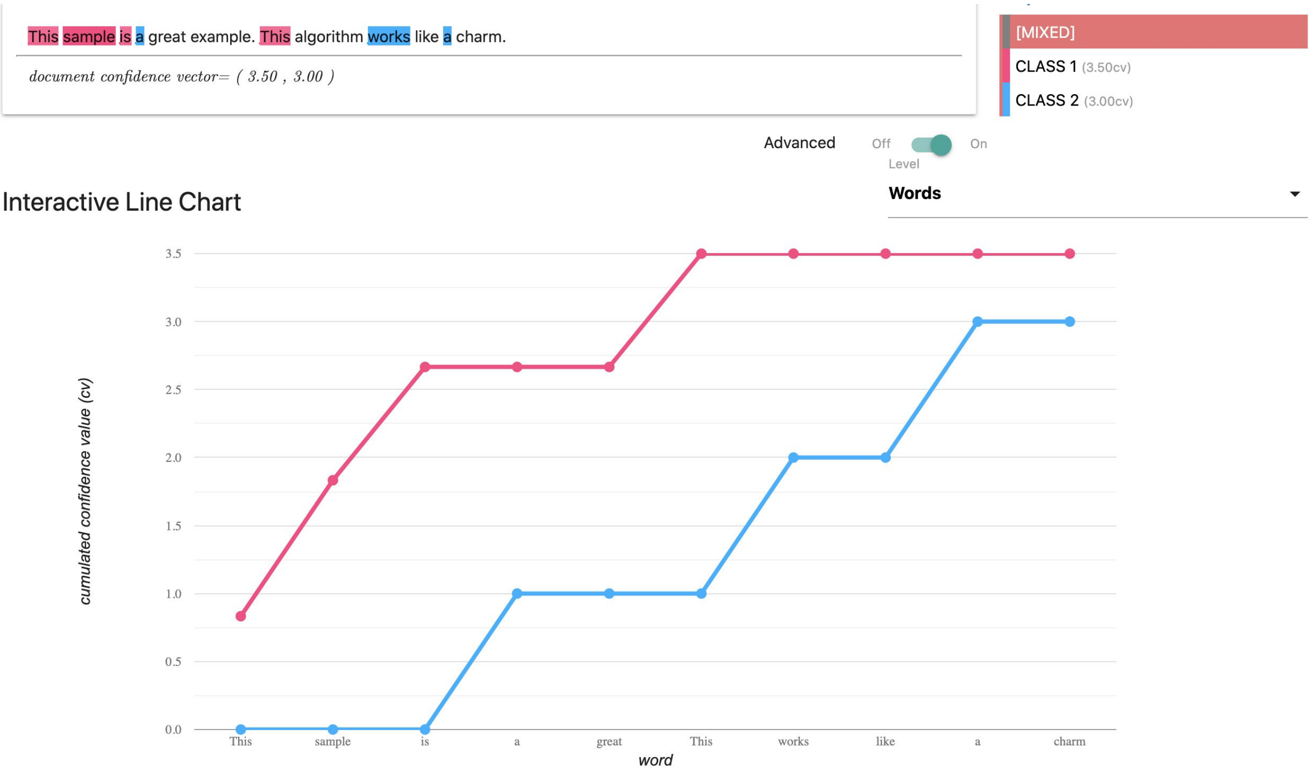
CV Class 1

Term	CV
sample	1.0
is	0.83
this	0.83
a	0
great	0
unique	0

CV Class 2

Term	CV
works	1.0
a	1.0
sample	0
like	0
charm	0
with	0

Explainability



Phase 1

SemEval 2016 Task 5

Dataset:

SemEval 2016 Task 5 Laptop & Restaurant Datasets

- No support for loading multi label data initially.
- Resolved this by duplicating test sample for each label.

Sample Text	Sample Label
Text 1	Label 1, Label 2
Text 2	Label 2

Sample Text	Sample Label
Text 1	Label 1
Text 1	Label 2
Text 2	Label 2

SemEval 2016 Task 5

```
<Reviews>
 <Review>
  <sentences>
   <sentence>
    <text>The wine list is interesting and has many good values.</text>
    <Opinions>
     <Opinion target="wine list" category="DRINKS#STYLE_OPTIONS" polarity="positive" from="4" to="13"/>
     <Opinion target="wine list" category="DRINKS#PRICES" polarity="positive" from="4" to="13"/>
    </Opinions>
   </sentence>
   <sentence>
    <text>For the price, you cannot eat this well in Manhattan.</text>
    <Opinions>
     <Opinion target="NULL" category="RESTAURANT#PRICES" polarity="positive" from="0" to="0"/>
     <Opinion target="NULL" category="FOOD#QUALITY" polarity="positive" from="0" to="0"/>
    </Opinions>
   </sentence>
  </sentences>
 </Review>
</Reviews>
```

Entity
Attribute

SemEval 2016 Task 5

Entity Dataset:

The wine list is interesting and has many good values.	DRINKS
For the price, you cannot eat this well in Manhattan.	RESTAURANT,FOOD

Attribute Dataset:

The wine list is interesting and has many good values.	STYLE_OPTIONS,PRICES
For the price, you cannot eat this well in Manhattan.	PRICES,QUALITY

Dataset Characteristics

Characteristics	Laptop	Restaurant
Multilabel	Yes	Yes
#Classes Entity	22	6
#Classes Attribute	9	5
Class Imbalance Entity (Max/Min)	1468/5	757/28
Class Imbalance Attribute (Max/Min)	804/51	998/97

Different Operators

“This” “sample” “is” “a” “great” “example”

(0.83,0) (1.0,0) (0.83,0) (0,1.0) (0,0) (0,0)

$\oplus$

(2.66, 1.0)

“This” “algorithm” “works” “like” “a” “charm”

(0.83, 0) (0,0) (0,0) (0,0) (0,1.0) (0,0)

$\oplus$

(0.83, 1.0)

$\oplus$

(3.49, 2.0)

Class 1

Different Operators

- SS3's default operator is sum.
- We changed the default operator to maximum and average and carried out the evaluation.
- Level 0 and Level 1 operator changed.

	Max (operator)		Avg (operator)	
Entity/Attribute	**Macro F1**	**Micro F1**	**Macro F1**	**Micro F1**
Entity	51 %	56 %	35 %	53 %
Attribute	57 %	53 %	51 %	59 %

Reduce Classes

- SS3 assign the most probable class in case it can’t classify, resulting in many false positive
- High imbalance, weak vocabulary for some classes impacts classifier performance
- What if we drop classes with low representation or merge classes with high overlap?

Entity/Attribute	Macro F1	Micro F1
Entity	46 %	49 %
Attribute	62 %	63 %

Bring Word Embeddings into the Game

Increase the power of classifier by adding words related to original sample

- Tokenize and remove stop words
- Calculate the x “closest” (euclidean distance) words and append to sample.
- Example:

  "I go to the restaurant." -> I go to the restaurant. go come gone restaurant cafe eatery

Results:

Attribute classification achieved **55% Macro F1**

Problems

- Words not in precalculated word embedding vector
- Problematic if class are really specific

Bigger Sample Size

- Above approaches did not improved much
- Changed sampling technique: sentence-wise → review-wise

Entity/Attribute	Macro F1	Micro F1
Entity	36 %	60%
Attribute	37 %	59 %

Laptop vs. Restaurant

	Laptop		Restaurant	
Attribute/Entity	Macro F1	Micro F1	Macro F1	Micro F1
Entity	42 %	47 %	59 %	64 %
Attribute	54 %	60 %	56 %	63 %

Phase 2

One vs. Rest Classification

- **Single Label:**
  Class with the highest confidence value is selected
- **Multi Label**:
  Specific class from each classifier (ignoring “others”)
  for example:
  Multi label classification with 3 classes (A, B, C)
  3 classifier each predicting either one of the 3 class or “others”:

Test	Classifier A (A or others)	Classifier B (B or others)	Classifier C (C or others)	Final Label
“This is a very good project”	A	B	others	A,B

One vs. Rest Classification

DATASET	One vs rest (Macro F1)	Default classifier (Macro F1)	One vs rest (Micro F1)	Default classifier (Micro F1)
Movie reviews	0.86	0.86	0.86	0.86
Beer review	0.60	0.60	0.70	0.69
SemEval2017	0.54	0.51	0.55	0.52
IMDB reviews	0.83	0.83	0.83	0.83
Restaurant	0.53	0.52	0.74	0.58
Movie Genre	0.54	0.50	0.56	0.51

Class Specific Hyperparameters

- CV:

$$cv(w,c)_\alpha = \begin{cases} 0 & lv_\sigma \cdot sg_\lambda \cdot sn_\rho < \alpha \\ lv_\sigma \cdot sg_\lambda \cdot sn_\rho & else \end{cases} \Rightarrow cv(w,c)_{\alpha_c} = \begin{cases} 0 & lv_{\sigma_c} \cdot sg_{\lambda_c} \cdot sn_{\rho_c} < \alpha_c \\ lv_{\sigma_c} \cdot sg_{\lambda_c} \cdot sn_{\rho_c} & else \end{cases}$$

- Local Value:

$$lv_\sigma(w,c) = (\frac{tf_{w,c}}{max\{tf_c\}})^\sigma \Rightarrow lv_{\sigma_c}(w,c) = (\frac{tf_{w,c}}{max\{tf_c\}})^{\sigma_c}$$

- Significance:

$$sg_\lambda(w,c) = \frac{1}{2}\tanh(4\frac{(lv(w,c) - \tilde{LV}_w)}{\lambda \cdot MAD_w} - 2) + \frac{1}{2} \Rightarrow sg_{\lambda_c}(w,c) = \frac{1}{2}\tanh(4\frac{(lv(w,c) - \tilde{LV}_w)}{\lambda_c \cdot MAD_w} - 2) + \frac{1}{2}$$

- Sanction:

$$sn_\rho(w,c) = (-\frac{\hat{C}_{w,c}}{\mid C \mid - 1} + 1)^\rho \Rightarrow sn_{\rho_c}(w,c) = (-\frac{\hat{C}_{w,c}}{\mid C \mid - 1} + 1)^{\rho_c}$$

Dataset Characteristics

Trivial

- Avg. #Words in Sample
- Avg. #Sentences in Sample
- Avg. #Paragraphs in Sample
- #Classes
- Imbalance: max(class)/min(class)

Dataset Characteristics

Avg. Important Word Ratio

- Ratio of important to unimportant words in dataset averaged over samples
- Important: Word that has cv > 0 for at least one class

Avg. Important Word Ratio Target

- Ratio of important to unimportant words in dataset averaged over classes
- Important: Word that has cv > 0 for target class

Word Overlap Ratio

- Ratio of words that occur in at least two classes

Important Word Overlap Ratio

- Ratio of important words that occur in at least two classes
- Important: Word that has cv > threshold

Overlap

$$overlap(target) = \frac{1}{len(words)} \sum_{word \in words} overlap(word, target)$$

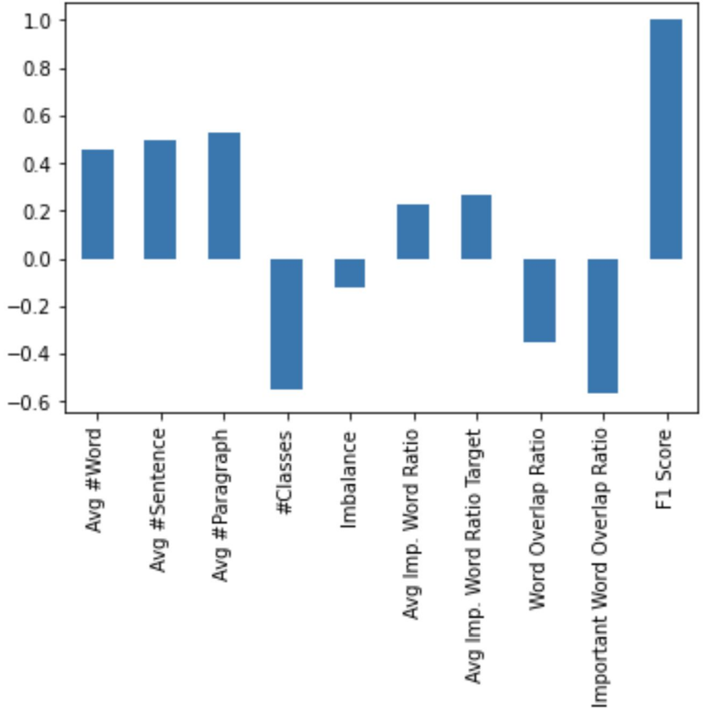
$$overlap(word, target) = \sum_{c \in classes/\{target\}} cv(c, word)$$

Taste: 0.06

Staff: 0.09

Pearson Correlation

```
F1 Score                             1.000000
Avg #Paragraph                       0.525862
Avg #Sentence                        0.497900
Avg #Word                            0.457430
Avg Imp. Word Ratio Target           0.269010
Avg Imp. Word Ratio                  0.222756
Imbalance                           -0.120185
Word Overlap Ratio                  -0.351182
#Classes                            -0.553842
Important Word Overlap Ratio        -0.568687
dtype: float64
```



	Avg #Word	Avg #Sentence	Avg #Paragraph	#Classes	Imbalance	Avg Imp. Word Ratio	Avg Imp. Word Ratio Target	Word Overlap Ratio	Important Word Overlap Ratio	F1 Score
Spam	18.892529	2.007853	1.000449	2.0	6.440735	0.442924	0.428792	0.148323	0.059447	0.921793
MovieReview	267.356007	12.552728	5.114754	2.0	1.575299	0.439056	0.437378	0.495548	0.433929	0.861986
IMDB	270.112381	12.513696	5.123446	2.0	1.015178	0.443151	0.442583	0.562934	0.315599	0.830358
TopicClasssification	28.458472	7.945024	1.000000	8.0	1.868425	0.537998	0.533980	0.302252	0.302000	0.694964
Apple	20.000767	1.562883	1.134969	3.0	5.350000	0.149534	0.138530	0.197989	0.197989	0.673001
Beer	6.638787	0.999865	1.000000	7.0	2.199052	0.791577	0.747992	0.458781	0.458781	0.623486
SemEval	23.801818	1.746259	1.000000	3.0	3.194036	0.136304	0.129037	0.450804	0.449308	0.517649
Alexa5	29.020238	2.300000	1.000000	5.0	23.740260	0.213439	0.196264	0.397746	0.397746	0.408180
Alexa2	29.118651	2.317063	1.000000	2.0	6.522388	0.384660	0.378119	0.347804	0.248004	0.408180
Clothing5	66.848361	4.645252	1.194965	5.0	15.449339	0.171963	0.167182	0.592775	0.592775	0.311794
Clothing2	66.910794	4.644241	1.201139	2.0	4.660741	0.432836	0.431334	0.463526	0.421737	0.311794
City	16.481412	1.079985	1.000000	24.0	179.2500	0.461586	0.405149	0.453799	0.453799	0.123188

Phase 3

Phase 3 - TODO: Decide which slide to use

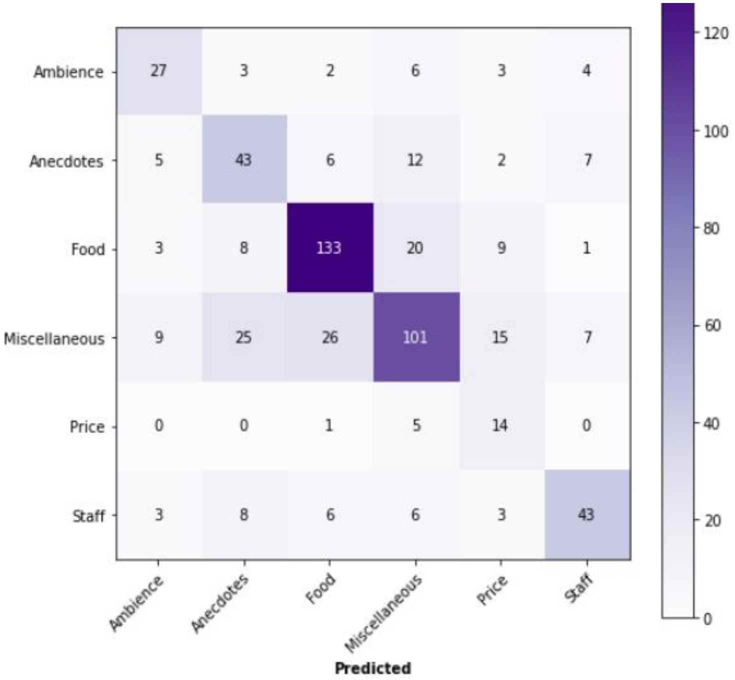
Why is beer performing better than city?

	precision	recall	f1-score	support
None	0.76	0.52	0.62	181
feel	0.83	0.82	0.83	205
look	0.97	0.87	0.92	323
overall	0.66	0.81	0.73	412
smell	0.70	0.83	0.76	270
taste	0.76	0.68	0.72	457
accuracy			0.76	1848
macro avg	0.78	0.76	0.76	1848
weighted avg	0.77	0.76	0.76	1848

	precision	recall	f1-score	support
Ambience	0.69	0.20	0.31	54
Ambience Anecdotes	0.00	0.00	0.00	1
Ambience Miscellaneous	0.00	0.00	0.00	1
Anecdotes	0.50	0.22	0.31	72
Anecdotes Miscellaneous	0.00	0.00	0.00	0
Food	0.72	0.50	0.59	172
Food Ambience	0.25	0.15	0.19	13
Food Anecdotes	0.00	0.00	0.00	5
Food Miscellaneous	0.00	0.00	0.00	7
Food Price	0.16	0.12	0.13	26
Food Staff	0.13	0.10	0.11	21
Miscellaneous	0.46	0.53	0.50	176
Neutral	0.00	0.00	0.00	1
Positive	0.00	0.00	0.00	1
Price	0.57	0.15	0.24	27
Price Ambience	0.00	0.00	0.00	3
Price Anecdotes	0.00	0.00	0.00	1
Price Miscellaneous	0.01	0.50	0.02	2
Price Staff	0.00	0.00	0.00	1
Staff	0.57	0.34	0.43	67
Staff Ambience	0.00	0.00	0.00	8
Staff Anecdotes	0.00	0.00	0.00	7
Staff Miscellaneous	0.00	0.00	0.00	0
accuracy			0.36	666
macro avg	0.18	0.12	0.12	666
weighted avg	0.51	0.36	0.41	666

City review with reduced classes:

	precision	recall	f1-score	support
Ambience	0.57	0.60	0.59	45
Anecdotes	0.49	0.57	0.53	75
Food	0.76	0.76	0.76	174
Miscellaneous	0.67	0.55	0.61	183
Price	0.30	0.70	0.42	20
Staff	0.69	0.62	0.66	69
accuracy			0.64	566
macro avg	0.58	0.64	0.59	566
weighted avg	0.66	0.64	0.64	566



Phase 3

Run classifier on major aspects for city review:

- "other aspects do not show clear patterns in either word usage or writing style.." [1]

Result?

	precision	recall	f1-score	support
Ambience	0.57	0.60	0.59	45
Anecdotes	0.49	0.57	0.53	75
Food	0.76	0.76	0.76	174
Miscellaneous	0.67	0.55	0.61	183
Price	0.30	0.70	0.42	20
Staff	0.69	0.62	0.66	69
accuracy			0.64	566
macro avg	0.58	0.64	0.59	566
weighted avg	0.66	0.64	0.64	566

[1] He, Ruidan, et al. "An unsupervised neural attention model for aspect extraction." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.

Vocabulary

	sentence	fn_word	cv	fn_class	count	cv_trueclass
185	in an effort to increase turnover , the resta...	no	0.020322	Food	1	0.024419
186	in an effort to increase turnover , the resta...	complimentary	0.039800	Food	1	0.060185
187	in an effort to increase turnover , the resta...	filled	0.039746	Food	1	0.000000
188	in an effort to increase turnover , the resta...	an	0.012790	Food	1	0.014410
189	in an effort to increase turnover , the resta...	chocolate	0.121635	Food	1	0.000000
190	in an effort to increase turnover , the resta...	desserts	0.054149	Food	1	0.000000
191	in an effort to increase turnover , the resta...	offers	0.039746	Food	1	0.050080
250	The service is not consistently excellent -- ...	decent	0.022638	Food	1	0.000000
251	The service is not consistently excellent -- ...	consistently	0.094314	Food	1	0.000000
252	The service is not consistently excellent -- ...	just	0.013194	Food	1	0.016623
253	The service is not consistently excellent -- ...	excellent	0.026819	Food	1	0.026587
254	Only drawback - they wo n't toast your bagel ...	n't	0.014643	Food	2	0.000000
255	Only drawback - they wo n't toast your bagel ...	bagel	0.036945	Food	2	0.000000
256	Only drawback - they wo n't toast your bagel ...	toast	0.069373	Food	1	0.000000
257	Only drawback - they wo n't toast your bagel ...	eggs	0.069373	Food	1	0.000000
258	Only drawback - they wo n't toast your bagel ...	make	0.025402	Food	1	0.030139
259	Only drawback - they wo n't toast your bagel ...	do	0.012154	Food	1	0.018921
260	Only drawback - they wo n't toast your bagel ...	your	0.011793	Food	1	0.015999
261	Only drawback - they wo n't toast your bagel ...	Only	0.011819	Food	1	0.013360
262	Only drawback - they wo n't toast your bagel ...	they	0.016829	Food	2	0.022941

Conclusion