Univariate Dictionary-based Classifiers BoP, BOSS, WEASEL

Erik Gottschalk, Xing Wang

SE "Landnutzungsklassifikation – als Wettbewerb" Humboldt-Universität zu Berlin

February 1, 2019

Structure

1. Introduction

2. WEASEL

- 2.1 Feature Selection
- 2.2 Fast Fourier Transformation (FFT)
- 2.3 F-Test for discriminative Fourier coefficients
- 2.4 Binning with Information Gain
- 2.5 Chi-Square for Dimensionality Reduction

3. Implementation

- 3.1 Preprocessing
- 3.2 Classification

4. Evaluation

- 4.1 Scalability
- 4.2 Per-Class Accuracy
- 4.3 Confusion Matrix
- 4.4 Common Errors

5. Conclusion

Structure

1. Introduction

2. WEASEL

- 2.1 Feature Selection
- 2.2 Fast Fourier Transformation (FFT)
- 2.3 F-Test for discriminative Fourier coefficients
- 2.4 Binning with Information Gain
- 2.5 Chi-Square for Dimensionality Reduction

3. Implementation

- 3.1 Preprocessing
- 3.2 Classification

4. Evaluation

- 4.1 Scalability
- 4.2 Per-Class Accuracy
- 4.3 Confusion Matrix
- 4.4 Common Errors

5. Conclusion

Bag of Patterns

BOSS

BOSS VS

WEASEL

• • • •

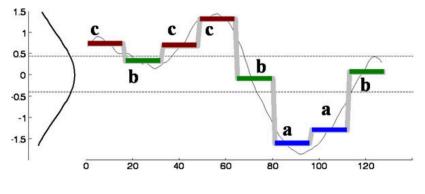


Figure: Bag of Patterns [1]

- Bag of Patterns
- BOSS
- BOSS VS
- WEASEL
- • •

Structure

1. Introduction

2. WEASEL

- 2.1 Feature Selection
- 2.2 Fast Fourier Transformation (FFT)
- 2.3 F-Test for discriminative Fourier coefficients
- 2.4 Binning with Information Gain
- 2.5 Chi-Square for Dimensionality Reduction

3. Implementation

- 3.1 Preprocessing
- 3.2 Classification

4. Evaluation

- 4.1 Scalability
- 4.2 Per-Class Accuracy
- 4.3 Confusion Matrix
- 4.4 Common Errors

5. Conclusion

- Data Preprocessing (NDVI)
- WEASEL Feature Selection/Transformation to Bag-of-Words
- Logistic Regression/SVM... Classifier

- Fast Fourier Transformation (FFT)
- F-Test for discriminative Fourier coefficients
- Binning with Information Gain (Bag-of-Words)
- Chi-Square for Dimensionality Reduction

- Input: Preprocessed Time Series NDVI
- Output: Fourier Coefficients

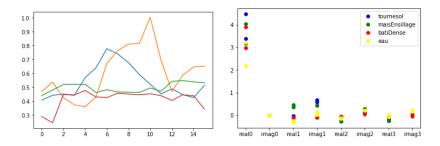


Figure: Example Time Series: 16 time steps

- Choose the f most discriminative Fourier coefficients
- f is the word length
- Example for *f* = 1 and one Fourier coefficient:

sample	class0	class1
0	0	2
1	3	5

- Choose the f most discriminative Fourier coefficients
- f is the word length
- Example for *f* = 1 and one Fourier coefficient:

sample	class0	class1
0	0	2
1	3	5
mean	1.5	3.5

Total mean: 2.5

F-Test for discriminative Fourier coefficients

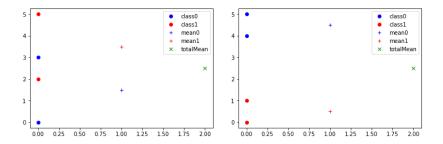


Figure: Plot of sample data: real0 (left), real1 (right)

F-Test for discriminative Fourier coefficients

$$F = \frac{MSB}{MSW}$$

$$MSB_{1} = (1.5 - 2.5)^{2} + (3.5 - 2.5)^{2} = 1$$

$$MSW_{1} = (0 - 1.5)^{2} + (3 - 1.5)^{2} + (2 - 3.5)^{2} + (5 - 3.5)^{2} = 4.5$$

$$F_{1} = \frac{1}{4.5} = 0.\overline{2}$$

$$MSB_{2} = (0.5 - 2.5)^{2} + (4.5 - 2.5)^{2} = 8$$

$$MSW_{2} = (0 - 0.5)^{2} + (1 - 0.5)^{2} + (4 - 4.5)^{2} + (5 - 4.5)^{2} = 1$$

$$F_{2} = \frac{8}{1} = 8$$

- Get the s most optimal split points for symbols
- s is the amount of the symbol alphabet
- Y_i : class labels

$$Ent(Y) = \sum_{(s_i, y_i) \in Y} -p_{y_i} \log_2 p_{y_i}$$
$$Ent(Y, x) = \frac{|Y_{Left}|}{|Y|} \cdot Ent(Y_{Left}) + \frac{|Y_{Right}|}{|Y|} \cdot Ent(Y_{Right})$$
$$InformationGain = Ent(Y) - Ent(Y, x)$$

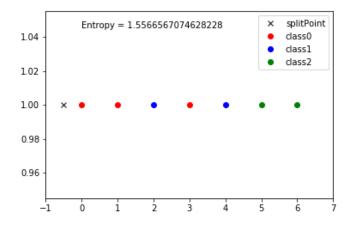


Figure: Fourier Coefficient real1: Find one split point, s = 2

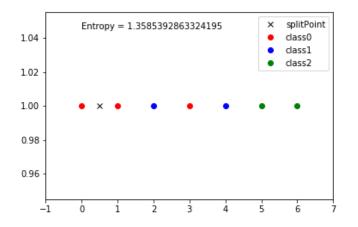


Figure: Fourier Coefficient real1: Find one split point, s = 2

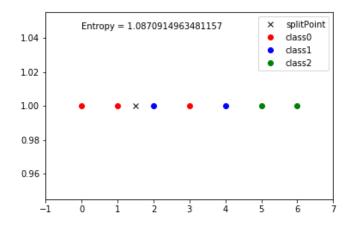


Figure: Fourier Coefficient real1: Find one split point, s = 2

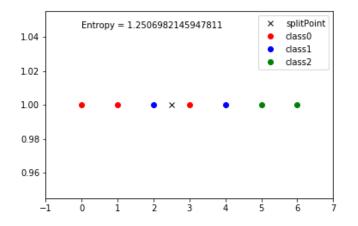


Figure: Fourier Coefficient real1: Find one split point, s = 2

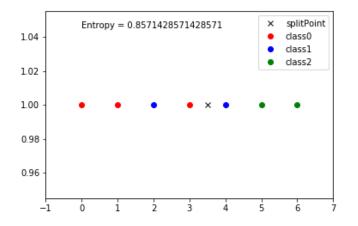


Figure: Fourier Coefficient real1: Find one split point, s = 2

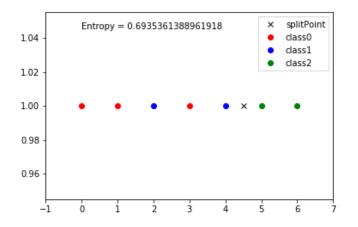


Figure: Fourier Coefficient real1: Find one split point, s = 2

Binning with Information Gain

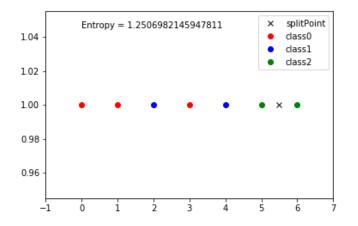


Figure: Fourier Coefficient real1: Find one split point, s = 2

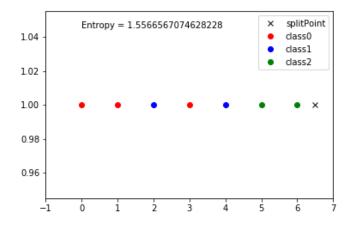


Figure: Fourier Coefficient real1: Find one split point, s = 2

Binning with Information Gain

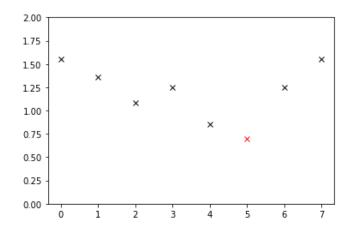


Figure: Fourier Coefficient real1: Max. Information Gain/Min. Entropy

Binning with Information Gain

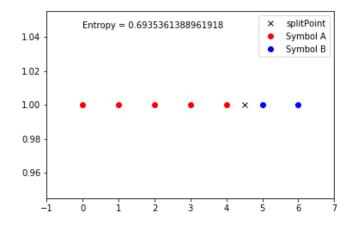


Figure: Fourier Coefficient real1: Splitting point into symbols 'A' & 'B'

- Dimensionality of Feature Space
- $\mathcal{O}(c^l)$ amount of words
- c is amount of symbols, I is word length
- using bigrams and $\mathcal{O}(n \cdot \log(n))$ non-overlapping windows yields $\rightarrow \mathcal{O}(c^{2l} \cdot n \cdot \log(n))$ features

Chi-Square for Dimensionality Reduction

	AA	AB	BA	BB	
1	350	50	50	0	450
2	150	50	50	50	300
3	100	0	0	100	200
4	0	0	0	50	50
	600	100	100	200	1000

Table: absolute frequency of words and classes

Chi-Square for Dimensionality Reduction

$$(80/270)^2 + (30/180)^2 + (20/120)^2 + (30/30)^2 < 2$$

AB and BA:

$$(5/45)^2 + (20/30)^2 + (20/20)^2 + (5/5)^2 > 2$$

BB:

$$(90/90)^2 + (10/60)^2 + (60/40)^2 + (40/10)^2 > 19$$

Structure

1. Introduction

2. WEASEL

- 2.1 Feature Selection
- 2.2 Fast Fourier Transformation (FFT)
- 2.3 F-Test for discriminative Fourier coefficients
- 2.4 Binning with Information Gain
- 2.5 Chi-Square for Dimensionality Reduction

3. Implementation

- 3.1 Preprocessing
- 3.2 Classification

4. Evaluation

- 4.1 Scalability
- 4.2 Per-Class Accuracy
- 4.3 Confusion Matrix
- 4.4 Common Errors

5. Conclusion

Preprocessing

Univariate \rightarrow NDVI

$$\mathsf{NDVI} = \frac{\mathsf{NIR} - \mathsf{Red}}{\mathsf{NIR} + \mathsf{Red}}$$

- NaNs → Interpolation: Python Pandas
 - Linear
 - Zero Filling
 - Quadratic Spline
 - Qubic Spline

Implementation

Preprocessing

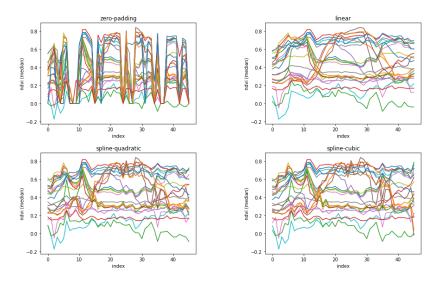


Figure: Imputation Methods; N = 100000

Implementation

Preprocessing

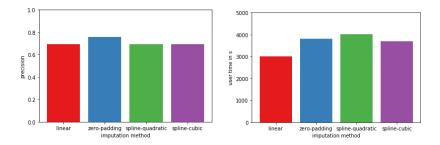


Figure: Influence of Imputation on Accuracy (left) and User Time (right), Train Set Size N = 100000

Classification

- Java Library for Time Series Data Analytics https://github.com/patrickzib/SFA
- Hyperparameters → Grid Search 90/10-Split
 - Min/Max Window Size
 - Max Alphabet Size

Implementation

Classification

minF	maxS											
=maxF	2	4	8	16								
2	0.583	0.579	0.618	0.632								
4	0.598	0.643	0.643	0.635								
6	0.624	0.654	0.645	0.640								
8	0.616	0.667	0.645	0.608								
16	0.641	0.636	0.624	0.060								
	2	4	8	16								
2	7071s	4769s	1247s	388s								
4	2075s	395s	337s	251s								
6	742s	282s	275s	296s								
8	472s	282s	320s	323s								
16	955s	691s	730s	130s								

Table: Top: Accuracy, Bottom: User Time – Train Set Size: N = 10000

Implementation

Classification

minF	maxS												
=maxF	2	4	8	16									
2													
4		0.662	0.675	0.668									
6		0.684	0.687	0.683									
8		0.689	0.694	0.680									
16		0.686	0.685	0.166									
	2	4	8	16									
2													
4		10920s	4535s	4197s									
6		5067s	3681s	3249s									
8		3804s	3015s	3112s									
16		3677s	3226s	846s									

Table: Top: Accuracy, Bottom: User Time – Train Set Size: N = 100000

Structure

1. Introduction

2. WEASEL

- 2.1 Feature Selection
- 2.2 Fast Fourier Transformation (FFT)
- 2.3 F-Test for discriminative Fourier coefficients
- 2.4 Binning with Information Gain
- 2.5 Chi-Square for Dimensionality Reduction

3. Implementation

- 3.1 Preprocessing
- 3.2 Classification

4. Evaluation

- 4.1 Scalability
- 4.2 Per-Class Accuracy
- 4.3 Confusion Matrix
- 4.4 Common Errors

5. Conclusion

Evaluation Scalability

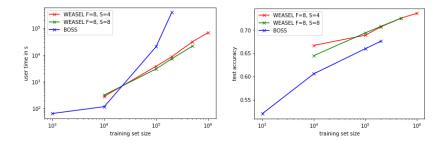


Figure: Influence of Train Set Size on User Time (left) and Accuracy (right)

Evaluation

Per-Class Accuracy

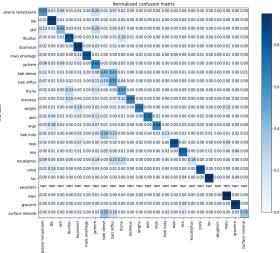
	precision		fl-score	
prairie temporaire	0.58	0.55	0.57	1629
ble	0.87	0.90	0.88	3291
pre	0.50	0.42	0.46	1252
feuillus	0.70	0.78	0.74	1250
tournesol	0.87	0.88	0.87	2060
mais ensillage	0.93	0.81	0.86	93
jachere	0.54	0.60	0.57	2725
bati dense	0.42	0.41	0.42	922
bati diffus	0.45	0.51	0.48	1592
friche	0.52	0.52	0.52	1180
resineux	0.83	0.63	0.71	121
sorgho	0.85	0.66	0.74	232
pois	0.97	0.71	0.82	133
orge	0.81	0.64	0.72	502
bati indu	0.43	0.13	0.20	154
soja	0.92	0.80	0.86	241
eau	0.79	0.73	0.76	147
eucalyptus	0.80	0.18	0.30	22
colza	0.89	0.81	0.85	449
lac	0.91	0.91	0.91	46
peupliers	0.00	0.00	0.00	Θ
mais	0.94	0.95	0.95	1784
graviere	0.85	0.91	0.88	113
surface mineral	0.56	0.35	0.44	62
weighted avg	0.70	0.69	0.69	20000

Evaluation **Confusion Matrix**

	Confusion matrix, without normalization																										
5	prairie temporaire	896	56	135	11	18	0	322	20	119	28	2	1	1	7	0	0	1	0	6	0	0	6	0	0	_	
	ble	42	2949	24	1	29	0	154	9	21	14	0	0	0	34	1	0	0	0	10	0	0	3	0	0		
	pre	159	23	532	34	9	0	314	12	65	91	0	0	0	5	0	0	0	0	5	0	0	3	0	0		
	feuillus	6	1	13	981	1	0	27	1	13	198	9	0	0	0	0	0	0	0	0	0	0	0	0	0		
	tournesol	31	10	9	7	1812	3	69	18	46	4	0	10	0	5	1	6	0	1	3	1	0	24	0	0		- 2500
	mais ensillage	0	0	0	0	10	75	1	0	0	0	0	1	0	0	0	4	0	0	1	0	0	1	0	0		
	jachere	218	145	217	47	67	1	1645	42	176	131	0	0	0	10	2	0	1	0	9	0	0	14	0	0		
	bati dense	17	11	10	2	7	0	71	377	394	7	0	0	0	1	9	0	5	0	0	0	0	1	1	9		
	bati diffus	104	18	44	26	18	0	169	310	810	67	2	2	0	2	7	0	5	0	0	0	0	3	1	4		- 2000
	friche	26	4	64	269	0	0	148	3	48	613	3	1	0	0	0	0	0	0	0	0	0	1	0	0		
	resineux	1	0	2	18	0	0	1	1	8	14	76	0	0	0	0	0	0	0	0	0	0	0	0	0		
bel	sorgho	1	2	1	0	42	0	5	1	2	1	0	152	0	0	0	1	1	0	0	0	0	23	0	0		- 1500
True label	pois	2	8	1	1	6	0	15	1	0	1	0	0	94	3	0	0	0	0	1	0	0	0	0	0		- 1500
¢=	orge	15	90	8	0	4	0	37	6	7	0	0	0	2	322	1	0	0	0	9	0	0	1	0	0		
	bati indu	7	4	0	0	8	0	7	55	36	0	0	0	0	4	20	0	3	0	1	1	0	1	3	4		
	soja	0	0	1	0	16	1	1	1	1	0	0	4	0	0	0	194	0	0	0	0	0	22	0	0		- 1000
	eau	0	1	0	2	2	0	3	8	9	1	0	0	0	0	3	0	107	0	0	2	0	0	9	0		1000
	eucalyptus	2	1	1	1	0	0	3	0	5	5	0	0	0	0	0	0	0	4	0	0	0	0	0	0		
	colza	5	57	1	0	2	0	13	0	6	0	0	0	0	3	0	0	0	0	362	0	0	0	0	0		
	lac	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	42	0	0	2	0		- 500
	peupliers -	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		500
	mais	6	4	1	0	36	1	17	1	8	2	0	8	0	0	0	6	0	0	0	0	0	1694	0	0		
	graviere -	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9	0	0	0	0	0	103	0		
	graviere -	0	0	0	0	1	0	1	22	11	0	0	0	0	0	1	0	2	0	0	0	0	0	2	22		L _o
	surface mineral	1,	-	-	-	-	-	-			-	-	-	-	-	-	-	-		-	-		-	-	-		0
		prairie temporaire	ble	pre	feuillus	tournesol	mais ensillage	jachere	bati dense	bati diffus.	friche	resineux	sorgho	pois		bati indu	soja	nea	eucalyptus	colza	lac	peupliers	mais	graviere	surface mineral		

Predicted label

Evaluation Confusion Matrix



Predicted label

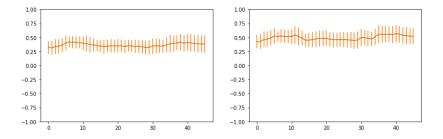


Figure: bati dense vs. bati diffuse

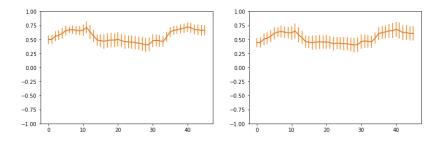


Figure: pre vs. jachere

Structure

1. Introduction

2. WEASEL

- 2.1 Feature Selection
- 2.2 Fast Fourier Transformation (FFT)
- 2.3 F-Test for discriminative Fourier coefficients
- 2.4 Binning with Information Gain
- 2.5 Chi-Square for Dimensionality Reduction

3. Implementation

- 3.1 Preprocessing
- 3.2 Classification

4. Evaluation

- 4.1 Scalability
- 4.2 Per-Class Accuracy
- 4.3 Confusion Matrix
- 4.4 Common Errors

5. Conclusion

- general method for time series classification
- few code adjustments
- lots of hyperparameters
- good scalability
- optimization: majority vote of multiple models (Ensemble)

References I

[1] Lin, J., Khade, R. & Li, Y.

Rotation-invariant similarity in time series using bag-of-patterns representation. Journal of Intelligent Information Systems, 39(2):287–315, 2012.

- [2] Schäfer, P. The BOSS is concerned with time series classification in the presence of noise. Data Mining and Knowledge Discovery, 29(6):1505–1530, 2015.
- Schäfer, P. & Leser, U.
 Fast and accurate time series classification with WEASEL.
 Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 1505–1530, 2015.