

SEMANTIC ANALYSIS FOR AUTOMATED EVALUATION OF THE POTENTIAL IMPACT OF RESEARCH ARTICLES

NESLIHAN SUZEN, ALEXANDER GORBAN, JEREMY LEVESLEY,
AND EVGENY MIRKES

1. THE SPACES USED IN CLASSIFICATION

¹SCHOOL OF MATHEMATICS AND ACTUARIAL SCIENCE, UNIVERSITY OF LEICESTER, LEICESTER
LE1 7RH, UK

²LOBACHEVSKY UNIVERSITY, NIZHNI NOVGOROD, RUSSIA

E-mail addresses: ns433@leicester.ac.uk, ns553@leicester.ac.uk (N. Suzen),
ag153@leicester.ac.uk (A.N. Gorban), jl1@leicester.ac.uk (J. Levesley),
em322@leicester.ac.uk (E.M. Mirkles).

TABLE 1.1. Dimensions of the FVT vectors in the original space and the reduced space constructed after applying supervised PCA for the category Mathematics, Applied

FVT	Original Space	H/L classes			EH/EL classes		
		Kaiser Rule	Broken Stick	Supervised PCA	Kaiser Rule	Broken Stick	Supervised PCA
$(\vec{\mu})$	252	36	15	14	32	15	14
$(\vec{PC_1})$	252	25	8	8	24	8	8
$(\vec{\mu}, \vec{PC_1})$	504	59	22	16	59	22	22
$(\vec{c_1}, \vec{c_2})$	504	41	20	19	41	20	14
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	756	62	24	23	63	24	24

TABLE 1.2. Dimensions of FVT vectors in the original space and the space constructed after applying supervised PCA for the category Biology

FVT	Original Space	H/L classes			EH/EL classes		
		Kaiser Rule	Broken Stick	Supervised PCA	Kaiser Rule	Broken Stick	Supervised PCA
$(\vec{\mu})$	252	35	13	12	36	14	12
$(\vec{PC_1})$	252	29	10	8	30	11	9
$(\vec{\mu}, \vec{PC_1})$	504	64	25	25	64	25	25
$(\vec{c_1}, \vec{c_2})$	504	50	23	23	49	22	20
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	756	76	30	30	76	30	30

TABLE 1.3. Dimensions of FVT vectors in the original space and the reduced space constructed after applying supervised PCA for the category Management

FVT	Original Space	H/L classes			EH/EL classes		
		Kaiser Rule	Broken Stick	Supervised PCA	Kaiser Rule	Broken Stick	Supervised PCA
$(\vec{\mu})$	252	36	13	11	36	13	13
$(\vec{PC_1})$	252	23	9	9	24	9	9
$(\vec{\mu}, \vec{PC_1})$	504	62	23	23	61	22	22
$(\vec{c_1}, \vec{c_2})$	504	39	16	11	40	16	11
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	756	61	25	25	62	25	25

2. RESULTS

e.

TABLE 2.1. Results of the citation classifier LDA according to H/L for the category Mathematics, Applied

FVT	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
$(\vec{\mu})$	67.91	49.61	117.52	77.02	29.14	106.16	68.28	44.55	112.83
$(\vec{PC_1})$	61.66	57.02	118.68	63.42	44.68	108.10	63.49	47.24	110.73
$(\vec{\mu}, \vec{PC_1})$	58.83	63.73	122.56	62.91	46.47	109.38	59.12	54.37	113.49
$(\vec{c_1}, \vec{c_2})$	67.40	49.53	116.93	49.54	58.89	108.43	57.93	50.82	108.75
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	64.15	58.85	123.00	55.37	54.96	110.33	54.13	59.47	113.61

TABLE 2.2. Results of the citation classifier LDA according to EH/EL for the category Mathematics, Applied

FVT	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
$(\vec{\mu})$	69.93	54.69	124.62	74.37	35.51	109.88	69.93	47.31	117.24
$(\vec{PC_1})$	65.20	58.90	124.11	69.82	39.70	109.52	66.68	47.34	114.03
$(\vec{\mu}, \vec{PC_1})$	71.56	58.37	129.93	68.60	44.69	113.30	65.08	52.61	117.68
$(\vec{c_1}, \vec{c_2})$	68.66	54.16	122.82	55.94	55.04	110.98	53.98	56.76	110.74
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	67.25	62.97	130.21	54.19	58.03	112.23	64.89	53.56	118.45

TABLE 2.3. Results of the citation classifier LDA with LOOCV in according to H/L for the category Mathematics, Applied (for the Original Space only)

FVT	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
$(\vec{\mu})$	65.24	48.29	113.53
$(\vec{PC_1})$	58.36	56.35	114.71
$(\vec{\mu}, \vec{PC_1})$	60.30	57.90	118.20
$(\vec{c_1}, \vec{c_2})$	62.87	48.25	111.12
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	59.00	56.00	115.5

TABLE 2.4. The best results for the kNN citation classifier for H/L classification for each FVT for the category Mathematics, Applied

FVT	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
$(\vec{\mu})$	11	62.60	48.16	110.76	13	65.53	41.97	107.50	17	68.14	41.06	109.19
$(\vec{PC_1})$	13	68.46	40.76	109.22	17	63.54	44.16	107.70	17	62.82	44.13	106.95
$(\vec{\mu}, \vec{PC_1})$	13	68.46	40.77	109.23	11	52.87	55.75	108.62	17	65.54	43.10	108.64
$(\vec{c_1}, \vec{c_2})$	11	57.30	50.07	107.38	13	63.50	43.06	106.56	13	61.87	43.68	105.55
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	11	56.25	53.20	109.45	17	66.78	41.02	107.80	17	65.96	42.77	108.73

TABLE 2.5. The best results for the kNN citation classifier for EH/EL classification for each FVT for the category Mathematics, Applied

FVT	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
$(\vec{\mu})$	11	77.83	37.99	115.82	11	68.41	44.57	112.98	17	70.53	45.94	116.48
$(\vec{PC_1})$	17	73.52	41.22	114.74	17	61.69	49.09	110.78	17	61.87	49.55	111.42
$(\vec{\mu}, \vec{PC_1})$	17	73.52	41.22	114.74	11	66.55	45.88	112.43	17	67.74	46.97	114.70
$(\vec{c_1}, \vec{c_2})$	13	72.21	39.80	112.01	17	65.92	44.40	110.32	17	60.13	47.77	107.90
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	17	73.63	40.93	114.56	17	66.22	45.31	111.53	17	67.81	46.37	114.18

TABLE 2.6. Results of the citation classifier LDA for H/L classification for the category Biology

FVT	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
$(\vec{\mu})$	74.30	69.21	143.51	64.29	63.48	127.77	70.47	64.33	134.80
$(\vec{PC_1})$	72.50	63.87	136.38	64.29	52.04	116.33	71.72	55.80	127.52
$(\vec{\mu}, \vec{PC_1})$	75.38	71.60	146.98	65.66	63.45	129.11	70.77	59.82	130.59
$(\vec{c_1}, \vec{c_2})$	70.70	69.12	139.82	67.53	50.45	117.98	58.89	58.90	117.79
$(\vec{c_1}, \vec{c_2}, \vec{PC_1})$	78.69	67.06	145.76	68.02	55.09	123.11	67.66	64.16	131.82

TABLE 2.7. Results of the citation classifier LDA for EH/EL classification for the category Biology

	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT									
$(\vec{\mu})$	79.91	79.44	159.34	67.18	69.70	136.87	71.05	75.14	146.20
(\vec{PC}_1)	75.87	72.37	148.24	69.57	53.03	122.60	71.76	65.95	137.70
$(\vec{\mu}, \vec{PC}_1)$	81.63	81.42	163.05	70.47	68.83	139.30	75.28	67.28	142.56
(\vec{c}_1, \vec{c}_2)	77.63	76.70	154.33	70.07	54.15	124.22	49.94	74.57	124.51
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	81.12	80.66	161.78	67.53	63.28	130.80	71.80	72.04	143.84

TABLE 2.8. The best results of the citation classifier kNN for each FVT for H/L classification for the category Biology

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT												
$(\vec{\mu})$	17	70.64	59.72	130.36	17	64.39	61.02	125.40	13	60.88	67.04	127.92
(\vec{PC}_1)	17	63.37	59.59	122.96	15	28.84	82.18	111.02	17	58.92	60.65	119.57
$(\vec{\mu}, \vec{PC}_1)$	17	63.34	59.59	122.93	17	64.91	61.99	126.90	17	62.03	60.01	122.04
(\vec{c}_1, \vec{c}_2)	11	65.60	58.36	123.96	17	62.19	59.49	121.68	17	61.11	58.95	120.06
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	17	63.99	59.56	123.55	17	61.73	59.46	121.19	17	60.72	60.71	121.43

TABLE 2.9. The best results of the citation classifier kNN for each FVT for EH/EL classification for the category Biology

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT												
$(\vec{\mu})$	17	82.33	60.61	142.94	17	73.21	64.83	138.03	17	74.19	66.31	140.49
(\vec{PC}_1)	17	72.42	61.00	133.43	17	66.51	61.69	128.20	17	66.98	64.54	131.52
$(\vec{\mu}, \vec{PC}_1)$	17	72.42	61.00	133.43	17	72.78	64.25	137.03	17	69.72	61.18	130.91
(\vec{c}_1, \vec{c}_2)	17	71.84	59.63	131.47	17	67.06	63.56	130.62	17	66.94	62.55	129.49
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	13	72.39	61.94	134.33	17	66.98	63.74	130.72	17	70.07	62.16	132.23

TABLE 2.10. Results of the LDA citation classifier for H/L classification for the category Management

	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT									
$(\vec{\mu})$	79.55	63.93	143.48	70.56	51.25	121.81	76.10	58.19	134.28
(\vec{PC}_1)	72.82	65.94	138.76	65.45	49.19	114.64	71.31	56.85	128.16
$(\vec{\mu}, \vec{PC}_1)$	80.05	65.82	145.87	69.20	53.18	122.39	70.20	62.49	132.69
(\vec{c}_1, \vec{c}_2)	79.55	60.96	140.51	48.67	69.12	117.79	53.23	62.47	115.70
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	79.89	65.85	145.74	56.87	61.36	118.23	78.38	54.54	132.92

TABLE 2.11. Results of the LDA citation classifier for EH/EL classification for the category Management

	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT									
$(\vec{\mu})$	85.20	77.46	162.66	72.24	60.17	132.40	80.30	68.86	149.16
(\vec{PC}_1)	80.68	74.21	154.88	66.86	54.21	121.07	70.38	70.00	140.38
$(\vec{\mu}, \vec{PC}_1)$	83.22	81.81	165.03	67.24	66.56	133.80	78.97	69.13	148.10
(\vec{c}_1, \vec{c}_2)	81.96	75.82	157.78	51.88	72.00	123.88	64.27	61.50	125.77
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	82.21	82.54	164.75	74.62	52.92	127.54	75.95	73.77	149.72

TABLE 2.12. The best results of the citation classifier kNN for each vector according to H/L for the category Management

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT												
$(\vec{\mu})$	17	74.40	52.10	126.50	17	65.09	52.89	117.98	17	67.67	57.80	125.47
(\vec{PC}_1)	17	71.83	49.12	120.94	17	61.59	55.39	116.98	17	62.34	57.62	119.96
$(\vec{\mu}, \vec{PC}_1)$	17	71.83	49.12	120.94	17	68.28	52.80	121.08	17	66.29	54.72	121.01
(\vec{c}_1, \vec{c}_2)	13	65.50	56.01	121.51	17	63.47	52.85	116.32	17	60.71	56.11	116.82
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	17	71.74	49.08	120.81	17	65.07	51.94	117.01	17	65.14	54.98	120.12

TABLE 2.13. The best results for the kNN citation classifier for each FVT for EH/EL classification for the category Management

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)	<i>k</i>	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT												
$(\vec{\mu})$	17	82.01	60.09	142.10	17	74.62	57.49	132.11	17	78.19	63.31	141.50
(\vec{PC}_1)	17	80.10	52.92	133.02	17	71.26	57.43	128.68	17	71.93	60.95	132.88
$(\vec{\mu}, \vec{PC}_1)$	17	80.10	52.94	133.04	17	76.23	57.07	133.30	17	76.26	57.39	133.65
(\vec{c}_1, \vec{c}_2)	7	69.62	63.72	133.35	17	73.39	56.25	129.64	11	72.89	55.79	128.68
$(\vec{c}_1, \vec{c}_2, \vec{PC}_1)$	7	69.55	62.74	132.28	17	74.25	55.26	129.51	11	65.55	68.02	133.58

TABLE 2.14. LOOCV results for the best classification performance among all categories: LDA with LOOCV for EH/EL classification for the Management category. The results correspond to the FVT (μ, PC_1) in the Original Space.

	<i>Sens.</i> (%)	<i>Spec.</i> (%)	<i>Sum</i> (%)
FVT			
$(\vec{\mu}, \vec{PC}_1)$	80.43	78.90	159.33

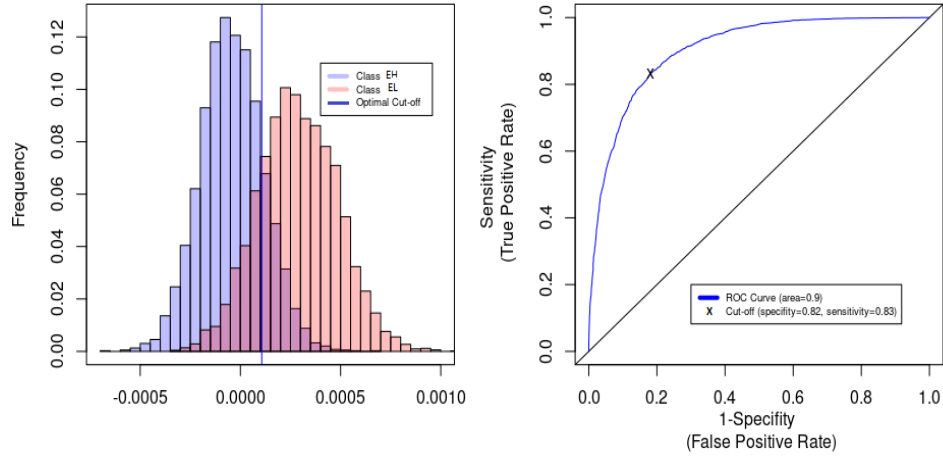


FIGURE 1. (Left) Data projection on the first LDA axis (Right) ROC curve for LDA. Both figures correspond to the best performance of classifiers: LDA for classes extremely little-cited (EL) and extremely highly-cited (EH) with the FVT (μ, PC_1) in the original space for the Management category.