

Supplemental Material: Multimodal Analytics for Real-world News using Measures of Cross-modal Entity Consistency

ERIC MÜLLER-BUDACK, TIB – Leibniz Information Centre for Science and Technology

JONAS THEINER, Leibniz Universität Hannover

SEBASTIAN DIERING, Leibniz Universität Hannover

MAXIMILIAN IDAHL, L3S Research Center, Leibniz Universität Hannover

RALPH EWERTH, L3S Research Center, Leibniz Universität Hannover and TIB – Leibniz Information Centre for Science and Technology

EVALUATION OF IMAGE SOURCES

This section supplements the evaluation of different images sources and number of images discussed in Section 4.3 of the paper. In total, we gathered a maximum of 20 images from the image search engines of *Google* and *Bing* as well as all k_W available images on *Wikidata* (mostly one *Wikimedia* image) for each entity recognized in the text. We have used multiple sources to prevent possible selection biases of a specific image source and investigated the performance for different images sources and number of images. Results for document verification with respect to verification accuracy (VA) and collection retrieval with respect to AUC are presented in Table 1 and Table 2. *Wikidata* usually only provides one or in many cases no image for many entities. Thus, a comparison is not reasonable. In the experiments the maximum operator was used to calculate the individual entity similarities (explained in Section 3.3 of the paper). For persons, the feature vector was first filtered using the agglomerative clustering approach presented in Section 3.3.1. As reported in the paper the results have demonstrated that the performance using a single or all image sources are very similar. In fact, results were sometimes a little worse when more images were used, which could indicate that the first images crawled from the search engines might contain less noise.

EVALUATION OF OPERATORS TO COMBINE THE CROSS-MODAL SIMILARITIES OF AN ENTITY

In Table 2 and Section 4.3 of the paper, different operators to combine the cross-modal similarities of all reference images to the news image for each entity (as explained in Section 3.3 of the paper) were evaluated. Results were solely reported with respect to collection retrieval using the AUC metric. In Table 3 the results for document verification with respect to the verification accuracy are reported and allow for similar conclusions.

RESULTS ON OTHER SUBSETS OF TAMPEREDNEWS AND NEWS400

Results for the top-25% documents according to the untampered cross-modal similarity values as well as for all documents of *TamperedNews* are presented in Table 4 and Table 5. As discussed in Section 4.3 of the paper we have used subsets of *TamperedNews* in order to counteract the influence of untampered documents that do not contain a single cross-modal entity relation. Thus, results for all documents are worse compared to the proposed subsets, since many documents without cross-modal relations are considered. Similar conclusion can be achieved for *TamperedNews* (Top-25%)

Authors' addresses: Eric Müller-Budack, TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany, eric.mueller@tib.eu; Jonas Theiner, Leibniz Universität Hannover, Hannover, Germany, theiner@stud.uni-hannover.de; Sebastian Diering, Leibniz Universität Hannover, Hannover, Germany, diering@stud.uni-hannover.de; Maximilian Idahl, L3S Research Center, Leibniz Universität Hannover, Hannover, Germany, idahl@l3s.de; Ralph Ewerth, L3S Research Center, Leibniz Universität Hannover, TIB – Leibniz Information Centre for Science and Technology, ralph.ewerth@tib.eu.

Table 1. Verification accuracy (VA) for different image sources (W - Wikidata, G - Google, B - Bing) using a maximum of I images on the *TamperedNews (Top-50%)* validation dataset. Results are reported for the hardest tampering strategy. Notations for tampering are according to Section 4.1.1 in the paper.

	I_W	I_G	I_B	Persons	Locations		Events
				PsCG	Outdoor GCD_{200}^{25}	Indoor GCD_{200}^{25}	EsP
Google	—	20	—	0.94	0.76	0.67	0.70
Bing	—	—	20	0.90	0.75	0.70	0.77
Combined-10	all	10	10	0.93	0.75	0.71	0.73
Combined-20	all	20	20	0.93	0.77	0.69	0.75

Table 2. AUC for different image sources (W - Wikidata, G - Google, B - Bing) using a maximum of I images on the *TamperedNews (Top-50%)* validation dataset. Results are reported for the hardest tampering strategy. Notations for tampering are according to Section 4.1.1 in the paper.

	I_W	I_G	I_B	Persons	Locations		Events
				PsCG	Outdoor GCD_{200}^{25}	Indoor GCD_{200}^{25}	EsP
Google	—	20	—	0.94	0.70	0.64	0.69
Bing	—	—	20	0.91	0.71	0.67	0.71
Combined-10	all	10	10	0.94	0.71	0.68	0.69
Combined-20	all	20	20	0.93	0.72	0.67	0.69

Table 3. Verification accuracy (VA) for different operators to calculate the cross-modal similarity for each entity of a given type (Section 3.3 of the paper) within a document. Results are reported for the *TamperedNews (Top-50%)* validation dataset with the hardest tampering strategy. Notations are according to Section 4.1.1 of the paper.

Test set	$ D $	clustering	Q_{75}	Q_{90}	Q_{95}	max
Persons: PsCG	1,703	0.93	0.92	0.94	0.93	0.90
Locations - Outdoor: GCD_{200}^{25}	1,420	—	0.72	0.73	0.74	0.75
Locations - Indoor: GCD_{200}^{25}	1,973	—	0.66	0.68	0.68	0.71
Events: EsP	839	—	0.72	0.73	0.73	0.73

and *TamperedNews (Top-50%)*. However, in particular results while retrieving tampered documents are noticeable better when using (smaller) subsets. As discussed in Section 4.4.2 this is mainly caused by the fact that some untampered entities in the documents that are depicted in both image and text, can be either unspecific (e.g. mentioning of a country) or the retrieved images for visual verification do not fit the document’s image content. When using these subsets, we have bypassed this problem. We have verified the same behavior for *News400* when experimenting on a subsets with the top-50% documents according to the untampered cross-modal similarity values (25% subset is omitted since it contains too few documents). The respective results are shown in Table 6.

Table 4. Results for document verification and collection retrieval for the *TamperedNews* (Top-25%) dataset for different entity testsets. Notations are according to Section 4.1.1 of the paper.

Test set ($ D $)	Document Verification Verification Accuracy	AUC	Collection Retrieval					
			AP-clean			AP-tampered		
			@25%	@50%	@100%	@25%	@50%	@100%
Persons (7,662)								
Random	0.98	0.98	96.63	96.58	96.35	100.0	100.0	98.74
PsC	0.98	0.98	96.28	96.27	95.93	100.0	100.0	98.56
PsG	0.98	0.98	96.14	96.14	96.05	100.0	100.0	98.68
PsCG	0.98	0.98	95.38	95.75	95.72	100.0	100.0	98.60
Locations 15,108)								
• Outdoor (6,390)								
Random	0.94	0.93	96.44	94.04	90.48	100.0	100.0	95.23
GCD(750, 2500)	0.93	0.90	92.32	89.62	85.96	100.0	100.0	93.02
GCD(200, 750)	0.87	0.85	89.23	85.13	80.07	100.0	100.0	88.89
GCD(25, 200)	0.85	0.82	86.69	82.29	77.14	100.0	100.0	86.73
• Indoor (8,718)								
Random	0.83	0.81	76.59	74.73	73.74	100.0	100.0	87.23
GCD(750, 2500)	0.81	0.78	67.90	68.11	68.78	100.0	100.0	84.64
GCD(200, 750)	0.82	0.79	73.21	72.01	71.20	100.0	100.0	85.49
GCD(25, 200)	0.76	0.75	59.04	61.85	64.56	100.0	100.0	82.67
Events (3,777)								
Random	0.95	0.94	93.40	92.62	90.90	100.0	100.0	96.16
EsP	0.81	0.77	74.41	71.86	69.72	100.0	100.0	83.16
Context (18,108)								
Random	0.92	0.92	94.24	92.41	88.83	100.0	100.0	94.48
Similar (top-25%)	0.90	0.89	89.48	87.55	84.27	100.0	100.0	92.38
Similar (top-10%)	0.87	0.85	84.32	82.08	79.35	100.0	100.0	89.88
Similar (top-5%)	0.84	0.82	80.36	78.09	75.66	100.0	100.0	87.68

Table 5. Results for document verification and collection retrieval for all documents in the *TamperedNews* dataset for different entity testsets. Notations are according to Section 4.1.1 of the paper.

Test set ($ D $)	Document Verification Verification Accuracy	AUC	Collection Retrieval					
			AP-clean			AP-tampered		
			@25%	@50%	@100%	@25%	@50%	@100%
Persons (30,646)								
Random	0.72	0.69	94.05	89.53	74.25	61.56	62.43	62.36
PsC	0.70	0.68	93.22	88.76	73.16	59.92	60.76	61.03
PsG	0.71	0.69	93.61	89.24	73.73	61.11	61.69	61.75
PsCG	0.70	0.68	93.31	88.71	72.88	59.01	59.82	60.43
Locations 60,433)								
• Outdoor (25,560)								
Random	0.67	0.64	84.89	76.94	66.18	60.64	60.59	59.03
GCD(750, 2500)	0.67	0.61	81.92	73.93	63.95	56.82	57.34	56.78
GCD(200, 750)	0.61	0.57	74.66	66.98	59.55	54.58	54.62	54.20
GCD(25, 200)	0.58	0.55	71.27	64.07	57.76	52.37	52.90	52.89
• Indoor (34,873)								
Random	0.54	0.54	60.37	57.21	54.13	53.81	53.17	52.44
GCD(750, 2500)	0.53	0.52	56.99	55.03	52.74	48.79	50.06	50.67
GCD(200, 750)	0.54	0.53	59.78	56.70	53.74	51.71	51.82	51.69
GCD(25, 200)	0.51	0.52	53.26	52.81	51.58	49.19	50.11	50.52
Events (15,107)								
Random	0.66	0.65	85.52	78.03	67.24	66.47	64.48	61.35
EsP	0.58	0.56	65.31	61.26	56.59	53.62	54.06	53.52
Context (72,433)								
Random	0.57	0.57	76.63	67.39	59.74	56.45	55.62	54.52
Similar (top-25%)	0.57	0.57	73.16	65.31	58.69	55.81	55.34	54.26
Similar (top-10%)	0.56	0.55	68.85	62.40	57.04	54.88	54.43	53.50
Similar (top-5%)	0.55	0.54	65.89	60.31	55.84	53.82	53.57	52.83

Table 6. Results for document verification and collection retrieval for the *News400* (*Top-50%*) dataset for different entity testsets. Notations are according to Section 4.1.1 of the paper.

Test set ($ D $)	Document Verification Verification Accuracy	AUC	Collection Retrieval					
			AP-clean			AP-tampered		
			@25%	@50%	@100%	@25%	@50%	@100%
Persons (58)								
Random	1.00	1.00	100.0	100.0	100.0	100.0	100.0	100.0
PsC	1.00	1.00	100.0	100.0	99.49	100.0	100.0	99.60
PsG	0.98	0.99	100.0	98.95	98.24	100.0	100.0	99.24
PsCG	1.00	1.00	100.0	100.0	99.82	100.0	100.0	99.84
Locations (35)								
• Outdoor (27)								
Random	1.00	0.99	100.0	100.0	99.44	100.0	100.0	99.52
GCD(750, 2500)	0.89	0.91	93.79	92.61	88.51	100.0	100.0	94.22
GCD(200, 750)	0.89	0.86	90.87	87.55	81.95	100.0	100.0	90.06
GCD(25, 200)	0.89	0.85	90.87	87.55	81.72	100.0	100.0	88.96
• Indoor (8)								
Random	1.00	0.89	100.0	95.00	88.35	100.0	100.0	92.41
GCD(750, 2500)	0.62	0.62	58.33	62.20	60.53	100.0	91.67	73.14
GCD(200, 750)	0.88	0.70	83.33	73.33	69.10	100.0	100.0	79.83
GCD(25, 200)	0.88	0.77	83.33	77.08	73.64	100.0	100.0	84.79
Events (16)								
Random	1.00	0.92	100.0	92.94	89.39	100.0	100.0	94.95
EsP	0.81	0.78	58.45	63.21	67.80	100.0	100.0	85.94
Context (46)								
Random	0.93	0.94	81.83	87.03	88.18	100.0	100.0	95.81
Similar (top-25%)	0.91	0.91	97.53	92.19	88.72	100.0	100.0	93.88
Similar (top-10%)	0.78	0.86	64.94	71.45	75.82	100.0	100.0	91.14
Similar (top-5%)	0.85	0.82	76.59	77.25	75.64	100.0	99.63	86.93