

Supplementary Material

February 23, 2018

1 Qualitative analysis of clusterings w.r.t. description size

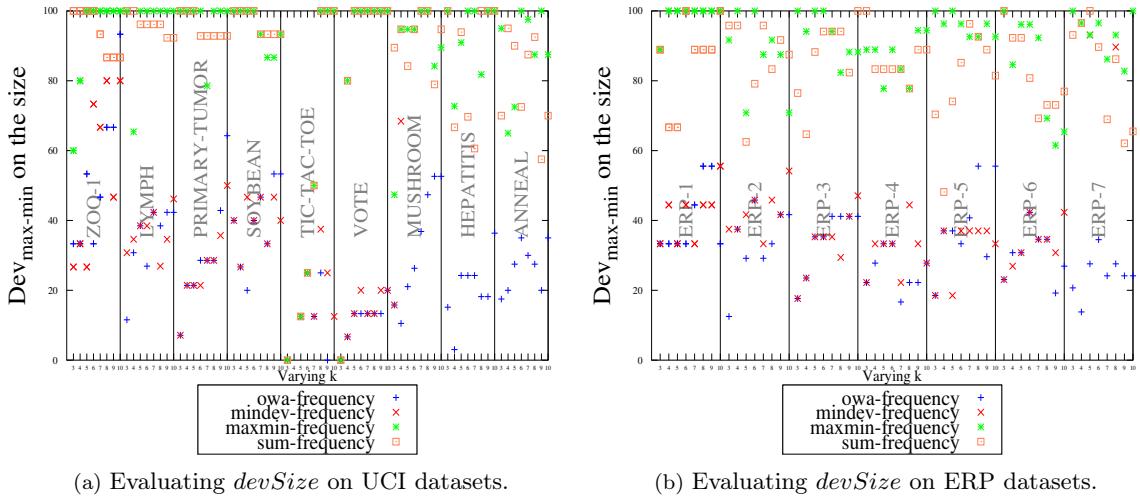


Figure 1: Quality of balancing of the resulting clusterings of the different ILP models w.r.t. description size.

2 Qualitative analysis of clusterings w.r.t. frequency

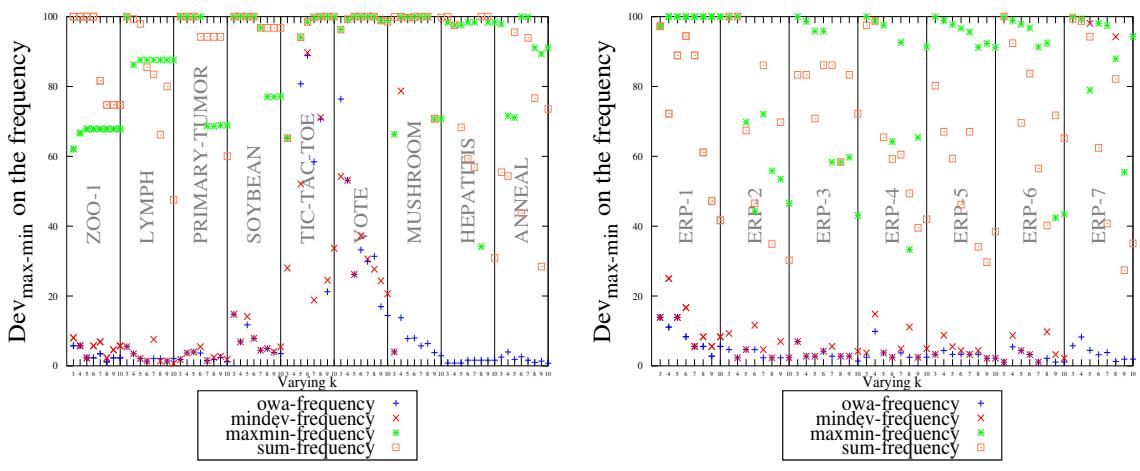


Figure 2: Quality of balancing of the resulting clusterings of the different ILP models w.r.t. frequency.

\mathcal{D}	k	OWA		minDev		maxMin		maxSum		\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD			ICS	ICD	ICS	ICD	ICS	ICD		
ERP1	3	0.554	0.779	0.554	0.779	1.000	0.380	0.554	0.779	ERP2	3	0.368	0.764	0.368	0.764	1.000	0.116	0.405	0.753
	4	0.400	0.859	0.400	0.859	0.998	0.380	0.427	0.844		4	0.284	0.849	0.277	0.850	0.963	0.173	0.284	0.849
	5	0.346	0.913	0.331	0.896	0.992	0.385	0.381	0.899		5	0.230	0.903	0.228	0.905	0.879	0.253	0.260	0.884
	6	0.285	0.945	0.285	0.945	0.979	0.389	0.302	0.941		6	0.185	0.933	0.209	0.927	0.722	0.391	0.230	0.909
	7	0.262	0.961	0.249	0.961	0.973	0.390	0.280	0.954		7	0.163	0.959	0.166	0.957	0.686	0.432	0.168	0.955
	8	0.209	0.982	0.214	0.980	0.970	0.391	0.251	0.966		8	0.135	0.978	0.145	0.974	0.517	0.579	0.143	0.970
	9	0.185	0.992	0.201	0.986	0.666	0.712	0.210	0.983		9	0.122	0.990	0.121	0.989	0.438	0.720	0.134	0.980
	10	0.171	1.000	0.170	0.999	0.663	0.713	0.177	0.996		10	0.106	1.000	0.108	0.998	0.431	0.720	0.119	0.994
(a) ERP-1										(b) ERP-2									
ERP3	3	0.377	0.773	0.377	0.773	1.000	0.076	0.454	0.678	ERP4	3	0.377	0.770	0.377	0.770	1.000	0.060	0.377	0.770
	4	0.291	0.854	0.291	0.853	0.973	0.104	0.375	0.787		4	0.310	0.845	0.299	0.849	0.983	0.096	0.304	0.835
	5	0.231	0.913	0.231	0.913	0.913	0.211	0.288	0.867		5	0.237	0.904	0.237	0.904	0.962	0.126	0.298	0.859
	6	0.197	0.941	0.216	0.931	0.781	0.376	0.270	0.882		6	0.196	0.939	0.200	0.935	0.911	0.166	0.209	0.931
	7	0.169	0.963	0.169	0.963	0.778	0.378	0.229	0.909		7	0.175	0.958	0.170	0.957	0.607	0.564	0.187	0.947
	8	0.146	0.979	0.145	0.977	0.630	0.526	0.181	0.952		8	0.151	0.973	0.152	0.973	0.883	0.182	0.209	0.937
	9	0.134	0.988	0.131	0.988	0.777	0.378	0.169	0.961		9	0.135	0.984	0.137	0.983	0.380	0.765	0.155	0.975
	10	0.115	0.999	0.115	1.000	0.732	0.426	0.154	0.969		10	0.119	1.000	0.122	0.996	0.185	0.951	0.145	0.987
(c) ERP-3										(d) ERP-4									
ERP5	3	0.363	0.766	0.363	0.766	1.000	0.064	0.588	0.494	ERP6	3	0.387	0.787	0.368	0.763	1.000	0.066	0.387	0.787
	4	0.283	0.844	0.277	0.850	0.979	0.088	0.354	0.752		4	0.284	0.855	0.297	0.870	0.971	0.072	0.356	0.831
	5	0.227	0.903	0.221	0.900	0.952	0.158	0.235	0.886		5	0.235	0.915	0.236	0.904	0.952	0.098	0.276	0.889
	6	0.190	0.937	0.187	0.936	0.914	0.207	0.190	0.937		6	0.202	0.947	0.198	0.947	0.337	0.804	0.272	0.904
	7	0.159	0.961	0.167	0.953	0.855	0.270	0.179	0.949		7	0.177	0.966	0.164	0.965	0.932	0.184	0.198	0.954
	8	0.151	0.975	0.151	0.975	0.872	0.245	0.162	0.968		8	0.148	0.977	0.149	0.981	0.864	0.228	0.195	0.957
	9	0.132	0.988	0.134	0.986	0.854	0.271	0.146	0.979		9	0.137	0.990	0.130	0.993	0.612	0.564	0.179	0.976
	10	0.119	0.998	0.116	1.000	0.226	0.906	0.129	0.990		10	0.122	1.000	—	—	0.346	0.815	0.165	0.984
(e) ERP-5										(f) ERP-6									
ERP7	3	0.375	0.778	—	—	1.000	0.030	0.375	0.778		3	0.375	0.787	0.368	0.763	1.000	0.066	0.387	0.787
	4	0.297	0.855	—	—	0.995	0.058	0.286	0.846		4	0.284	0.855	0.297	0.870	0.971	0.072	0.356	0.831
	5	0.240	0.903	—	—	0.965	0.074	0.288	0.869		5	0.235	0.915	0.236	0.904	0.952	0.098	0.276	0.889
	6	0.198	0.944	—	—	0.960	0.103	0.267	0.902		6	0.202	0.947	0.198	0.947	0.337	0.804	0.272	0.904
	7	0.167	0.956	—	—	—	—	0.225	0.931		7	0.177	0.966	0.164	0.965	0.932	0.184	0.198	0.954
	8	0.143	0.974	—	—	0.948	0.116	0.202	0.944		8	0.148	0.977	0.149	0.981	0.864	0.228	0.195	0.957
	9	0.131	0.990	—	—	—	—	0.206	0.945		9	0.137	0.990	0.130	0.993	0.612	0.564	0.179	0.976
	10	0.120	1.000	—	—	—	—	0.190	0.956		10	0.122	1.000	—	—	0.346	0.815	0.165	0.984
(g) ERP-7																			

Table 1: Comparing the quality of the resulting clusterings in terms of ICS and ICD on ERP datasets when maximizing the frequency.

3 ICS/ICD analysis w.r.t. frequency

\mathcal{D}	k	OWA		minDev		maxMin		maxSum		\mathcal{D}	k	OWA		minDev		maxMin		maxSum			
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD			ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD		
zoo-1	3	0.533	0.785	0.588	0.819	0.746	0.657	1.000	0.259	vote	3	0.803	0.628	0.698	0.675	1.000	0.612	1.000	0.612		
	4	0.398	0.854	0.413	0.864	0.743	0.658	0.998	0.262		4	0.544	0.782	0.544	0.782	1.000	0.612	1.000	0.612		
	5	0.396	0.951	0.373	0.935	0.742	0.659	0.993	0.262		5	0.445	0.897	0.445	0.897	1.000	0.612	1.000	0.612		
	6	0.304	0.952	0.316	0.958	0.741	0.659	0.987	0.267		6	0.429	0.927	0.436	0.914	1.000	0.612	1.000	0.612		
	7	0.261	0.969	0.277	0.975	0.741	0.659	0.740	0.539		7	0.328	0.945	0.428	0.947	1.000	0.612	1.000	0.613		
	8	0.229	0.982	0.231	0.983	0.711	0.675	0.640	0.632		8	0.324	0.964	0.416	0.972	0.999	0.613	0.999	0.613		
	9	0.206	0.993	0.199	0.986	0.689	0.682	0.643	0.633		9	0.251	0.983	0.304	0.978	0.790	0.633	0.761	0.637		
	10	0.185	1.000	0.180	0.993	0.635	0.693	0.641	0.634		10	0.229	1.000	0.273	0.990	0.791	0.632	0.756	0.638		
	(a) Zoo-1										(b) Vote										
	tic-tac-toe	3	1.000	0.719	0.924	0.765	1.000	0.719	1.000		0.719	lymph	3	0.373	0.757	0.382	0.765	1.000	0.040	1.000	0.040
5		0.807	0.806	0.693	0.873	0.961	0.738	0.961	0.738	4	0.298		0.859	0.294	0.854	0.805	0.299	0.993	0.061		
6		0.886	0.779	0.892	0.776	0.960	0.738	0.960	0.738	5	0.234		0.902	0.240	0.908	0.804	0.299	0.978	0.099		
7		0.594	0.919	0.471	0.972	0.960	0.738	0.960	0.738	6	0.194		0.933	0.197	0.935	0.804	0.299	0.791	0.347		
8		0.601	0.915	0.676	0.880	0.960	0.738	0.960	0.738	7	0.171		0.959	0.180	0.959	0.804	0.300	0.748	0.363		
9		0.408	1.000	0.409	0.993	0.960	0.738	0.960	0.738	8	0.153		0.979	0.151	0.977	0.804	0.300	0.497	0.604		
10		0.413	0.998	0.413	0.998	0.960	0.738	0.960	0.738	9	0.134		0.990	0.134	0.990	0.801	0.302	0.715	0.463		
(c) Tic-tac-toe										(d) Lymph											
mushroom		3	0.450	0.767	0.450	0.767	0.638	0.519	1.000	0.060	hepatitis		3	0.432	0.764	—	—	0.957	0.230	1.000	0.217
		4	0.347	0.833	0.708	0.439	1.000	0.060	1.000	0.060			4	0.331	0.852	—	—	0.952	0.236	0.953	0.237
	5	0.283	0.890	—	—	0.999	0.060	0.998	0.065	5		0.266	0.901	—	—	0.951	0.238	0.575	0.615		
	6	0.250	0.922	—	—	0.999	0.060	0.999	0.063	6		0.237	0.946	—	—	0.951	0.238	0.491	0.726		
	7	0.245	0.964	—	—	1.000	0.060	1.000	0.060	7		0.200	0.964	—	—	0.950	0.238	0.455	0.740		
	8	0.225	0.982	—	—	1.000	0.060	1.000	0.060	8		0.172	0.977	—	—	0.317	0.879	0.991	0.222		
	9	0.193	0.989	—	—	0.639	0.517	0.617	0.539	9		0.154	0.991	—	—	0.949	0.239	0.991	0.223		
	10	0.178	1.000	—	—	0.639	0.517	0.998	0.065	10		0.137	1.000	—	—	0.950	0.239	0.272	0.894		
	(e) Mushroom											(f) Hepatitis									
	Anneal	3	0.382	0.789	—	—	0.967	0.074	0.550	0.614		Primary-tumor	3	0.392	0.773	0.392	0.773	1.000	0.055	1.000	0.055
4		0.288	0.860	—	—	0.612	0.523	0.555	0.617	4	0.300		0.854	0.300	0.854	1.000	0.055	1.000	0.055		
5		0.237	0.914	—	—	0.595	0.551	0.932	0.133	5	0.248		0.907	0.248	0.907	1.000	0.055	1.000	0.055		
6		0.198	0.939	—	—	1.000	0.018	0.381	0.780	6	0.208		0.936	0.214	0.939	1.000	0.055	0.906	0.183		
7		0.174	0.966	—	—	0.998	0.022	0.898	0.165	7	0.182		0.962	0.182	0.962	0.556	0.583	0.906	0.183		
8		0.155	0.983	—	—	0.852	0.222	0.637	0.487	8	0.161		0.978	0.161	0.977	0.556	0.583	0.905	0.184		
9		0.135	0.988	—	—	0.811	0.243	0.242	0.898	9	0.143		0.989	0.145	0.990	0.556	0.583	0.904	0.185		
10		0.124	1.000	—	—	0.851	0.222	0.593	0.544	10	0.131		1.000	0.131	1.000	0.556	0.583	0.491	0.682		
(g) Anneal										(h) Primary-tumor											

Table 2: Comparing the quality of the resulting clusterings in terms of ICS and ICD on UCI datasets when maximizing the frequency.

\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD
ERP1	3	0.554	0.779	0.554	0.779	1.000	0.380	0.554	0.779
	4	0.400	0.859	0.400	0.859	0.998	0.380	0.427	0.844
	5	0.346	0.913	0.331	0.896	0.992	0.385	0.381	0.899
	6	0.285	0.945	0.285	0.945	0.979	0.389	0.302	0.941
	7	0.262	0.961	0.249	0.961	0.973	0.390	0.280	0.954
	8	0.209	0.982	0.214	0.980	0.970	0.391	0.251	0.966
	9	0.185	0.992	0.201	0.986	0.666	0.712	0.210	0.983
	10	0.171	1.000	0.170	0.999	0.663	0.713	0.177	0.996
	(a) ERP-1	(b) ERP-2							
	ERP3	3	0.377	0.773	0.377	0.773	1.000	0.076	0.454
4		0.291	0.854	0.291	0.853	0.973	0.104	0.375	0.787
5		0.231	0.913	0.231	0.913	0.913	0.211	0.288	0.867
6		0.197	0.941	0.216	0.931	0.781	0.376	0.270	0.882
7		0.169	0.963	0.169	0.963	0.778	0.378	0.229	0.909
8		0.146	0.979	0.145	0.977	0.630	0.526	0.181	0.952
9		0.134	0.988	0.131	0.988	0.777	0.378	0.169	0.961
10		0.115	0.999	0.115	1.000	0.732	0.426	0.154	0.969
(c) ERP-3		(d) ERP-4							
ERP5		3	0.363	0.766	0.363	0.766	1.000	0.064	0.588
	4	0.283	0.844	0.277	0.850	0.979	0.088	0.354	0.752
	5	0.227	0.903	0.221	0.900	0.952	0.158	0.235	0.886
	6	0.190	0.937	0.187	0.936	0.914	0.207	0.190	0.937
	7	0.159	0.961	0.167	0.953	0.855	0.270	0.179	0.949
	8	0.151	0.975	0.151	0.975	0.872	0.245	0.162	0.968
	9	0.132	0.988	0.134	0.986	0.854	0.271	0.146	0.979
	10	0.119	0.998	0.116	1.000	0.226	0.906	0.129	0.990
	(e) ERP-5	(f) ERP-6							
	ERP7	3	0.375	0.778			1.000	0.030	0.375
4		0.297	0.855			0.995	0.058	0.286	0.846
5		0.240	0.903			0.965	0.074	0.288	0.869
6		0.198	0.944			0.960	0.103	0.267	0.902
7		0.167	0.956					0.225	0.931
8		0.143	0.974			0.948	0.116	0.202	0.944
9		0.131	0.990					0.206	0.945
10		0.120	1.000					0.190	0.956
(g) ERP-7									

Table 3: Comparing the quality of the resulting clusterings in terms of ICS and ICD on ERP datasets when minimizing the diversity.

\mathcal{D}	k	OWA		minDev		maxMin		maxSum		\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD			ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD
zoo-1	3	0.531	0.808	0.549	0.816	1.000	0.176	0.584	0.799	vote	3	0.803	0.628	0.698	0.675	1.000	0.612	1.000	0.612
	4	0.428	0.885	0.428	0.885	0.999	0.177	0.473	0.884		4	0.544	0.782	0.544	0.782	1.000	0.612	0.598	0.789
	5	0.358	0.936	0.377	0.934	0.998	0.178	0.449	0.907		5	0.445	0.897	0.445	0.897	1.000	0.612	0.445	0.897
	6	0.333	0.951	0.339	0.948	0.996	0.179	0.419	0.932		6	0.419	0.919	0.522	0.914	1.000	0.612	0.680	0.878
	7	0.283	0.971	0.312	0.963	0.995	0.179	0.359	0.952		7	0.328	0.945	0.463	0.949	1.000	0.613	0.572	0.913
	8	0.261	0.984	0.264	0.979	0.785	0.388	0.338	0.966		8	0.324	0.964	0.923	0.751	0.999	0.613	0.514	0.955
	9	0.233	0.991	0.237	0.988	0.588	0.620	0.306	0.980		9	0.251	0.983	0.633	0.867	0.769	0.636	0.479	0.970
	10	0.207	1.000	0.229	0.998	0.773	0.395	0.276	0.987		10	0.229	1.000	0.829	0.734	0.794	0.632	0.417	0.987

(a) Zoo-1

(b) Vote

\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD
tic-tac-toe	3	0.924	0.763	0.924	0.763	1.000	0.717	1.000	0.717
	5	0.700	0.868	0.693	0.870	0.961	0.736	0.664	0.882
	6	0.886	0.777	0.892	0.774	0.960	0.736	0.886	0.777
	7	0.594	0.917	0.471	0.969	0.960	0.736	0.471	0.969
	8	0.601	0.913	0.676	0.878	0.960	0.736	0.601	0.913
	9	0.460	0.968	0.402	1.000	0.960	0.736	0.423	0.987
	10	0.532	0.936	0.413	0.996	0.960	0.736	0.429	0.989

(c) Tic-tac-toe

\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD
lymph	3	0.409	0.768	0.409	0.768	1.000	0.040	0.534	0.662
	4	0.302	0.858	0.315	0.850	0.993	0.061	0.404	0.790
	5	0.254	0.906	0.252	0.904	0.978	0.099	0.272	0.893
	6	0.225	0.931	0.234	0.917	0.958	0.134	0.248	0.915
	7	0.181	0.960	0.187	0.957	0.765	0.434	0.203	0.948
	8	0.157	0.977	0.163	0.971	0.893	0.251	0.194	0.956
	9	0.146	0.988	0.200	0.942	0.879	0.262	0.183	0.968
	10	0.126	1.000	0.143	0.989	0.879	0.263	0.232	0.936

(d) Lymph

\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD
mushroom	3	0.525	0.648	1.000	0.060	1.000	0.060	0.457	0.748
	4	0.417	0.798	1.000	0.060	1.000	0.060	0.417	0.798
	5	0.365	0.819	—	—	0.999	0.060	0.399	0.813
	6	0.323	0.937	—	—	0.999	0.060	0.323	0.937
	7	0.279	0.957	—	—	1.000	0.060	0.313	0.950
	8	0.261	0.976	—	—	1.000	0.060	0.290	0.970
	9	0.249	0.985	—	—	1.000	0.060	0.279	0.982
	10	0.207	1.000	—	—	1.000	0.060	0.280	0.982

(e) Mushroom

\mathcal{D}	k	OWA		minDev		maxMin		maxSum	
		ICS	ICD	ICS	ICD	ICS	ICD	ICS	ICD
hepatitis	3	1.000	0.787	—	—	—	—	—	—
	4	0.770	0.874	—	—	—	—	—	—
	5	0.648	0.930	—	—	—	—	—	—
	6	0.535	0.961	—	—	—	—	—	—
	7	0.471	0.984	—	—	—	—	—	—
	8	0.415	1.000	—	—	—	—	—	—

(f) Hepatitis

(g) Anneal

(h) Primary-tumor.

Table 4: Comparing the quality of the resulting clusterings in terms of ICS and ICD on UCI datasets when minimizing the diversity.