Gorilla–Sea Cucumber Hash

Dennis Sadeler Shapira & Jonas Lomholdt

April 9, 2015

Results

The following table gives the similarity between each pair of species as a number between 0 and 1, higher values meaning more similar. We have used the hash function that is built in to Java with $d = 10000$ and k-grams of length $k = 20$. As can be seen, the species closest to us is the *gorilla*.

	Human	Gorilla	Monkey	Horse	Deer	Pig	Cow	Gull	Trout	R. Cod	Lamprey	Sea Cuc.
Human	1.000	0.883	-0.247	-0.003	0.023	0.320	-0.051	-0.557	0.271	-0.218	0.287	-0.177
Gorilla	0.883	1.000	-0.499	-0.185	0.151	0.405	-0.163	-0.363	0.313	0.030	0.047	-0.196
Monkey	-0.247	-0.499	1.000	0.664	-0.673	0.133	0.201	-0.372	-0.572	-0.545	0.194	0.440
Horse	-0.003	-0.185	0.664	1.000	-0.233	0.275	0.338	-0.347	-0.368	-0.850	0.161	0.469
Deer	0.023	0.151	-0.673	-0.233	1.000	-0.378	-0.259	0.558	0.525	0.378	-0.181	-0.347
Pig	0.320	0.405	0.133	0.275	-0.378	1.000	0.314	-0.605	-0.340	-0.246	0.302	0.286
Cow	-0.051	-0.163	0.201	0.338	-0.259	0.314	1.000	-0.641	-0.781	-0.641	0.817	-0.403
Gull	-0.557	-0.363	-0.372	-0.347	0.558	-0.605	-0.641	1.000	0.609	0.647	-0.784	0.197
Trout	0.271	0.313	-0.572	-0.368	0.525	-0.340	-0.781	0.609	1.000	0.509	-0.630	0.231
R. Cod	-0.218	0.030	-0.545	-0.850	0.378	-0.246	-0.641	0.647	0.509	1.000	-0.481	-0.115
Lamprey	0.287	0.047	0.194	0.161	-0.181	0.302	0.817	-0.784	-0.630	-0.481	1.000	-0.556
Sea Cuc.	-0.177	-0.196	0.440	0.469	-0.347	0.286	-0.403	0.197	0.231	-0.115	-0.556	1.000

Tests

Our static method `double getSimilarity(Species a, Species b)` computes the cosine of the angle of two vectors of the same length d, from a Species object. Similarly, our static method `length(int[] p)` computes the vector length. We have tested it on the following examples:

p	q	d	value returned
$(0.0, 1.0)$	$(0.0, 1.0)$	2	1.0
$(0.0, 1.0)$	$(0.0, 2.0)$	2	1.0
$(0.0, 1.0)$	$(1.0, 0.0)$	2	0.0
$(0.0, 1.0)$	$(0.0, -1.0)$	2	-1.0
$(0.0, \ldots, 0.0, 1.0)$	$(1.0, 0.0, \ldots, 0.0)$	1000	0.7531

q	d	value returned
$(0.0, 1.0)$	2	1.0
$(0.0, 2.0)$	2	1.0
$(1.0, 0.0)$	2	2.0
$(0.0, -1.0)$	2	1.0
$(1.0, 0.0, \ldots, 0.0)$	1000	18.2950