

糖尿病罹患リスクを予測するヘルスケアデータの匿名化コンテスト PWS Cup 2021 データの解析

Analysis on Competition on Anonymized Healthcare Data to be qualified for Diabetes Prevalence Study, PWS Cup 2021

菊池 浩明*
Hiroaki Kikuchi

馬 瑞強*
MA Ruiqiang

キーワード ヘルスケア, 匿名化, 再識別

あらまし

件や課題を探る.

表 1: コンテスト結果

攻撃者	01	02	03	04	05	06	07	08	09	10	11	12	13	14	A	攻撃
01		0.520	0.005	0.016	0.010	0.005	0.002	0.025	0.006	0.045	0.486	0.192	0.035	0.026	0.022	6
02	0.068		0.003	0.003	0.002	0.003	0.006	0.005	0.004	0.042	0.004	0.002	0.003	0.007	0.016	9
03	0.093	0.970		0.004	0.049	0.001	0.000	0.004	0.006	0.034	0.462	0.003	0.008	0.119	0.013	12
04	0.010	0.007	0.011		0.002	0.024	0.002	0.035	0.026	0.030	0.018	0.002	0.020	0.056	0.027	5
05	0.005	0.005	0.005	0.005		0.067	0.015	0.283	0.257	0.057	0.005	0.005	0.046	0.036	0.043	2
06	0.022	0.015	0.005	0.015	0.026		0.008	0.045	0.017	0.067	0.037	0.005	0.057	0.057	0.041	3
07	0.054	0.721	0.003	0.005	0.028	0.001		0.004	0.005	0.036	0.236	0.013	0.009	0.009	0.014	11
08	0.027	0.026	0.005	0.015	0.057	0.015	0.005		0.109	0.026	0.005	0.005	0.036	0.015	0.019	7
09	0.081	0.046	0.005	0.067	0.165	0.036	0.011	0.067		0.057	0.048	0.057	0.077	0.036	0.053	1
10	0.027	0.026	0.005	0.026	0.015	0.046	0.026	0.221	0.166		0.027	0.036	0.036	0.057	0.036	4
11	0.059	0.857	0.003	0.008	0.018	0.002	0.002	0.003	0.005	0.024		0.002	0.002	0.007	0.011	13
12	0.042	0.008	0.006	0.016	0.048	0.008	0.008	0.002	0.014	0.007	0.008		0.010	0.012	0.010	14
13	0.011	0.007	0.003	0.004	0.021	0.031	0.008	0.039	0.016	0.012	0.005	0.007		0.000	0.016	10
14	0.023	0.203	0.003	0.011	0.007	0.001	0.000	0.002	0.018	0.043	0.084	0.002	0.009		0.018	8
R 参考	0.093	0.970	0.011	0.067	0.165	0.067	0.026	0.283	0.257	0.067	0.486	0.192	0.077	0.119	0.024	
R	0.093	0.970		0.067	0.165	0.067		0.283	0.257	0.067	0.486	0.192	0.077	0.119	1.000	
lmark	0.045	0.585	0.003	0.003	0.019	0.001	0.000	0.004	0.007	0.034	0.148	0.003	0.002	0.031		
加工	5	12		1	7	1		10	9	1	11	8	4	6		

健康診断やウェアラブルデバイスから取得したヘルスケアデータは生活習慣病の予測などに活用できる有益なビッグデータである。個人情報取扱事業者は、規則に従った適切な匿名加工に加えて、各種分析の精度を劣化させない最適な加工をすることが求められている。そこで、米国疾病対策予防センター CDC が収集した米国国民健康栄養調査 (National Health and Nutrition Examination Survey: NHANES) データを用いて、年齢、学歴、BMI、運動量などの説明変数に対する糖尿病の罹患リスクを正しく評価するための匿名化技術と再識別リスクを探求するコンテストを行った [1]。本稿では、本コンテストの結果を報告し、ヘルスケアデータの匿名化における陽性条

参考文献

- [1] 菊池, 他, PWS Cup 2021 -- 糖尿病罹患リスクを予測するヘルスケアデータの匿名化コンテスト, 情報処理学会コンピュータセキュリティシンポジウム (CSS) 2021.

* 明治大学総合数理学部, 東京都中野区中野 4-21-1 School of Interdisciplinary Mathematical Sciences, Meiji University, 4-21-1 Nakano, Nakano Ku, Tokyo, 164-8525 Japan, (kikn@meiji.ac.jp)