Binge drinking risk factors ranking using multi-task learning

Daniel Gallegos[1], Elaine Raymond[2], Lu Wang[3]

Abstract

Binge drinking is tied to increased severity of anxiety and depression symptoms.[1] Existing methods of identifying binge drinking risk factors often fail to account for subpopulation heterogeneity. We formulate a subpopulation binge drinking risk factors ranking problem, under the framework of multi-task learning (MTL), to obtain a ranked list of risk factors for each subpopulation (task).

Background

One in six US adults binge drinks, with one-fourth doing so at least weekly.[2] CDC tracks binge drinking and other health related risk behaviors in their Behavioral Risk Factor Surveillance System (BRFSS). The 2021 BRFSS had 438,693 records and 301 features, including the binge drinking calculated variable, _RFBING5. Binge drinkers are defined as males who consume 5+ drinks on one occasion, or females who consume 4+ drinks.[3]

Methods

Data preprocessing:
- Eliminated BRFSS entries with missing values
- 109,192 instances; 80 predictor variables; 1 output variable (_RFBING5 or _MENT14D)

Why MTL instead of Single-Task Learning (STL):
- STL uses one model to rank risk factors for all subpopulations
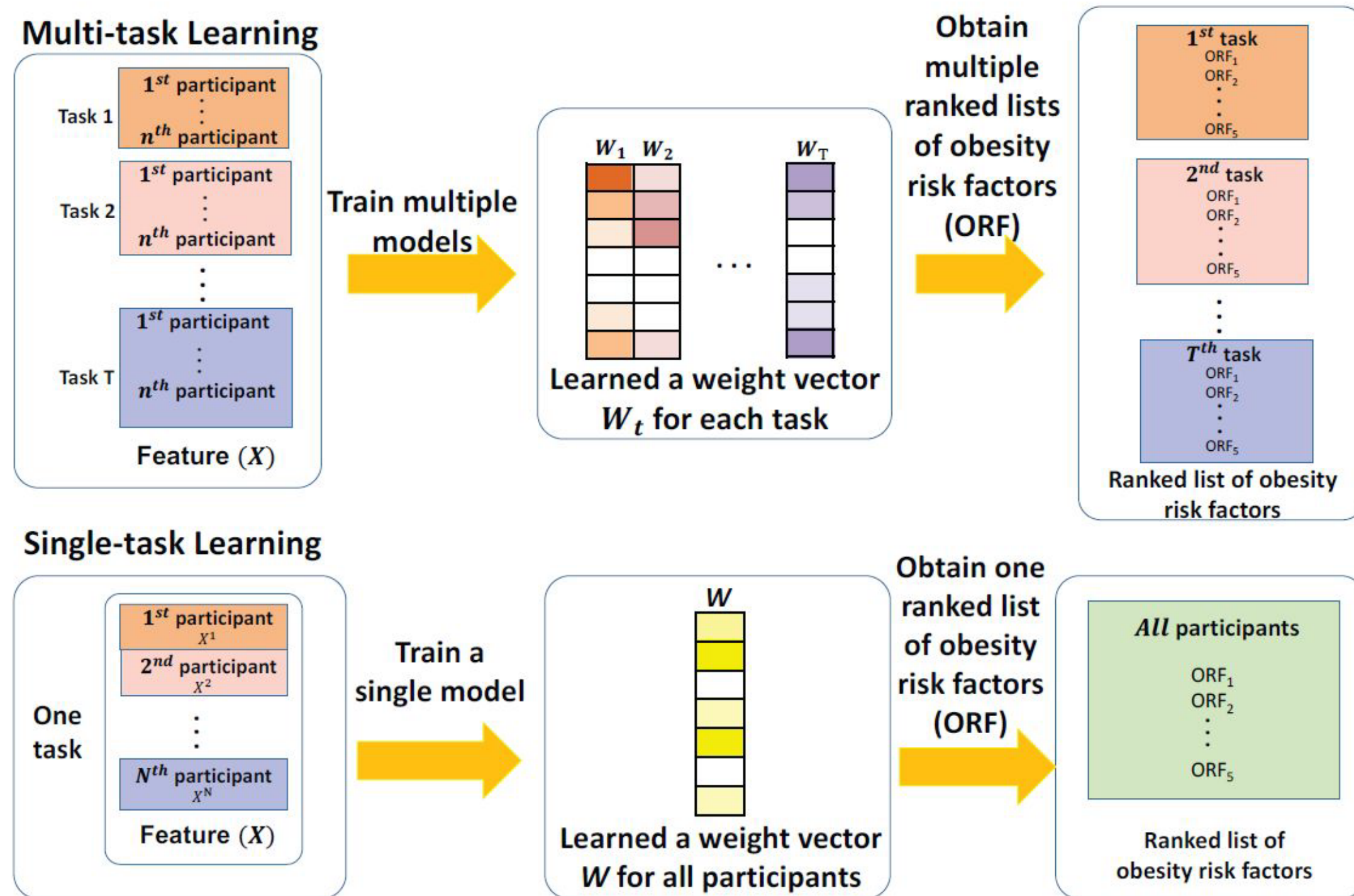- MTL trains multiple models simultaneously to obtain multiple ranked lists:[4]

Figure 1. MTL vs STL

Task Definition:
- Task setting 1 (MTL1) - 53 geographic information tasks (49 states and 4 territories)
- Task setting 2 (MTL2) - 3 mental health tasks defined by CDC: 0 days of not good mental health in the last 30 days; 1-13 days; 14+ days
- Task setting 3 (MTL3) - 2 binge drinking tasks defined by CDC: Non-binge drinkers with no drink and binge drinker with at least one drink multiple times

MALSAR: MTL via Structural Regularization MATLAB Package:
- Used l2,1-Norm Regularization with Least Squares Loss (Least L21) function
- Solves the l2,1-norm (and the squared l2-norm) regularized multi-task least squares problem:

$$\min_{W} \sum_{i=1}^{t} \|W_i^T X_i - Y_i\|_F^2 + \rho_1 \|W\|_{2,1} + \rho_{L2} \|W\|_F^2$$

Results and Conclusion

Figure 2. Top 5 Selected Risk Factors for MTL1

Figure 3. Top 10 Selected Risk Factors for MTL2

Figure 2 presents the top five ranking binge drinking risk factors for each geographic location in MTL1. The first column presents the variable name and description of each risk factor, and the colored cells include an abbreviation of each location.

Figure 3 presents the top ten ranking binge drinking risk factors for each mental health status in MTL2. Mental health status is abbreviated as MH1 for 0 days; MH2 for 1-13 days; MH3 for 14+ days. The first column presents the variable name and description of each risk factor.

Figure 4 presents the top ten ranking health risk factors for each binge drinking status in MTL3. Binge drinking status is abbreviated as BD1 for non-binge drinkers who have not had a drink in 30 days; BD2 for binge drinkers who consumed at least one drink on multiple occasions within 30 days.

Table 1 is a correlation matrix between the output variables and the geographic locations using Pearson's correlation.

Table 1. Pearson Correlation Matrix for binge drinking and mental health status and geographical location

	_MENT14D	_RFBING5	_STATE
_MENT14D	1.000000	0.014568	0.000445
_RFBING5	0.014568	1.000000	0.009416
_STATE	0.000445	0.009416	1.000000

Figure 4. Top 10 Selected Risk Factors for MTL3

Conclusion: MTL can be effectively utilized to obtain ranked lists of risk factors for binge drinking in various subpopulations. Several highly ranking risk factors were shared by most tasks across all task settings, including _AGE80 and _INCOMG1. Further work could compare these results with STL to determine if those features can be considered population-level risk factors.

Acknowledgements

References

[1] Lannoy, S., Duka, T., Carbia, C., Billieux, J., Fontesse, S., Dormal, V., ... & Maurage, P. (2021). Emotional processes in binge drinking: A systematic review and perspective. Clinical Psychology Review, 84, 101971.
[2] Centers for Disease Control and Prevention. (2022, November 14). Binge drinking. Centers for Disease Control and Prevention. https://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm
[3] Centers for Disease Control and Prevention. (2022, December 7). CDC - 2021 BRFSS survey data and Documentation. Centers for Disease Control and Prevention. https://www.cdc.gov/brfss/annual_data/annual_2021.html
[4] Wang, L., Zhu, D., Towner, E., & Dong, M. (2018, March). Obesity risk factors ranking using multi-task learning. In 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (pp. 385-388). IEEE.

[1] Western Governors University, dgal135@wgu.edu; [2] University of Houston, evraymond7@gmail.com; [3] Texas State University, lu.wang@txstate.edu