

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

Results and Conclusion

Five Year Age Group Between 18-80 (<code>_AGE80</code>)	AL AK AZ AR CA CO CT DE DC GA HI ID IL IN IA KS KY LA ME MD MA MI MN MS MO MT NE NV NH NJ NM NY NC ND OH OK OR PA RI SC SD TN TX UT VT VA WA WV WI WY GU PR VI
Drink Occasions Per Day (<code>DROCDY3_</code>)	AL AK AZ AR CA CO CT DE DC GA HI ID IL IN IA KS KY LA MA MI MN MS MO MT NE NV NJ NM NC ND OH OK OR PA RI SC SD TN TX UT VT WV WI WY GU PR VI
Highest Level of Education Completed (<code>EDUCA</code>)	AL AK AZ AR CA CO CT DE DC GA ID IL IN IA KS KY LA ME MD MA MI MN MS MO MT NE NV NH NJ NM NY NC ND OH OK OR PA RI SC SD TN UT VT VA WA WI
Income Category (<code>_INCOMG1</code>)	AK AR CA CO CT DE DC GA ID IN IA KS KY ME MD MA MI MN MO MT NE NV NH NJ NY NC ND OH OR PA RI SC SD UT VT VA WA WV WI WY
Primary Source of Health Insurance (<code>PRIMINSR</code>)	AL AK AZ CO DC ID KY LA ME MS MO MT NJ NM NC OK PA SD UT VT WI WY GU
Children in Household (<18) (<code>CHILDREN</code>)	AL DC IL LA NV OK OR TN PR VI
Computed Number of Children (<code>_CHLDCNT</code>)	IN IA KS ME NE NH NY ND WA WV
Computed Education Level (<code>_EDUCAG</code>)	CT GA MD MI RI SC VA
Computed Race Group (<code>_LRACEPRV</code>)	AZ CA HI NM TX
Last Time Cholesterol Checked by Doctor (<code>CHOLCHK3</code>)	DE MN NH NY OH
Four-Level Smoker Status (<code>_SMOKER3</code>)	MD MA VA
Self-Reported General Health (<code>GENHLTH</code>)	AR MS WA
Number of Days with Not Good Physical Health (<code>PHYSHLTH</code>)	GU PR VI
Number of Days with Not Good Mental Health (<code>MENTHLTH</code>)	PR VI
Calculated Multiracial Race Categorization (<code>_MRACE1</code>)	GU
Fourteen-Level Age Category (<code>_AGEG5YR</code>)	IL
Imputed Race/Ethnicity Value (<code>_IMPRACE</code>)	HI
Race/Ethnicity Categories (<code>_LRACE</code>)	HI

Number of Days with Not Good Mental Health (<code>MENTHLTH</code>)	MH1 MH2 MH3	Computed Level of Education Completed (<code>_EDUCAG</code>)	MH2
Told by Doctor of High Blood Pressure (<code>CHOLCHK3</code>)	MH1 MH2 MH3	Consumed fruit 1 or more times per day (<code>_FRTL1A</code>)	MH1
Computed Income Category (<code>_INCOMG1</code>)	MH2 MH3	Drink Occasions Per Day (<code>DROCDY3_</code>)	MH3
Computed Number of Children in Household (<code>_CHLDCNT</code>)	MH2 MH3	Dual Phone Use Categories (<code>_LDUALUSE</code>)	MH1
Five Year Age Group Between 18-80 (<code>_AGE80</code>)	MH2 MH3	Ever Told of Diabetes (<code>DIABETE4</code>)	MH2
Level of Education Completed (<code>_EDUCA</code>)	MH2 MH3	Heavy Alcohol Consumption (<code>_RFRHV7</code>)	MH1
Self-Reported General Health (<code>GENHLTH</code>)	MH1 MH3	Imputed Race/Ethnicity Value (<code>_IMPRACE</code>)	MH1
US State or Territory (<code>_STATE</code>)	MH2 MH3	Marital Status (<code>MARITAL</code>)	MH1
Children in Household (<18) (<code>CHILDREN</code>)	MH1	Primary Source of Health Insurance (<code>PRIMINSR</code>)	MH3
Computed Race Group (<code>_LRACEPRV</code>)	MH1	Smoked At Least 100 Cigarettes in Life (<code>USENOW3</code>)	MH2

Figure 3. Top 10 Selected Risk Factors for MTL2

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

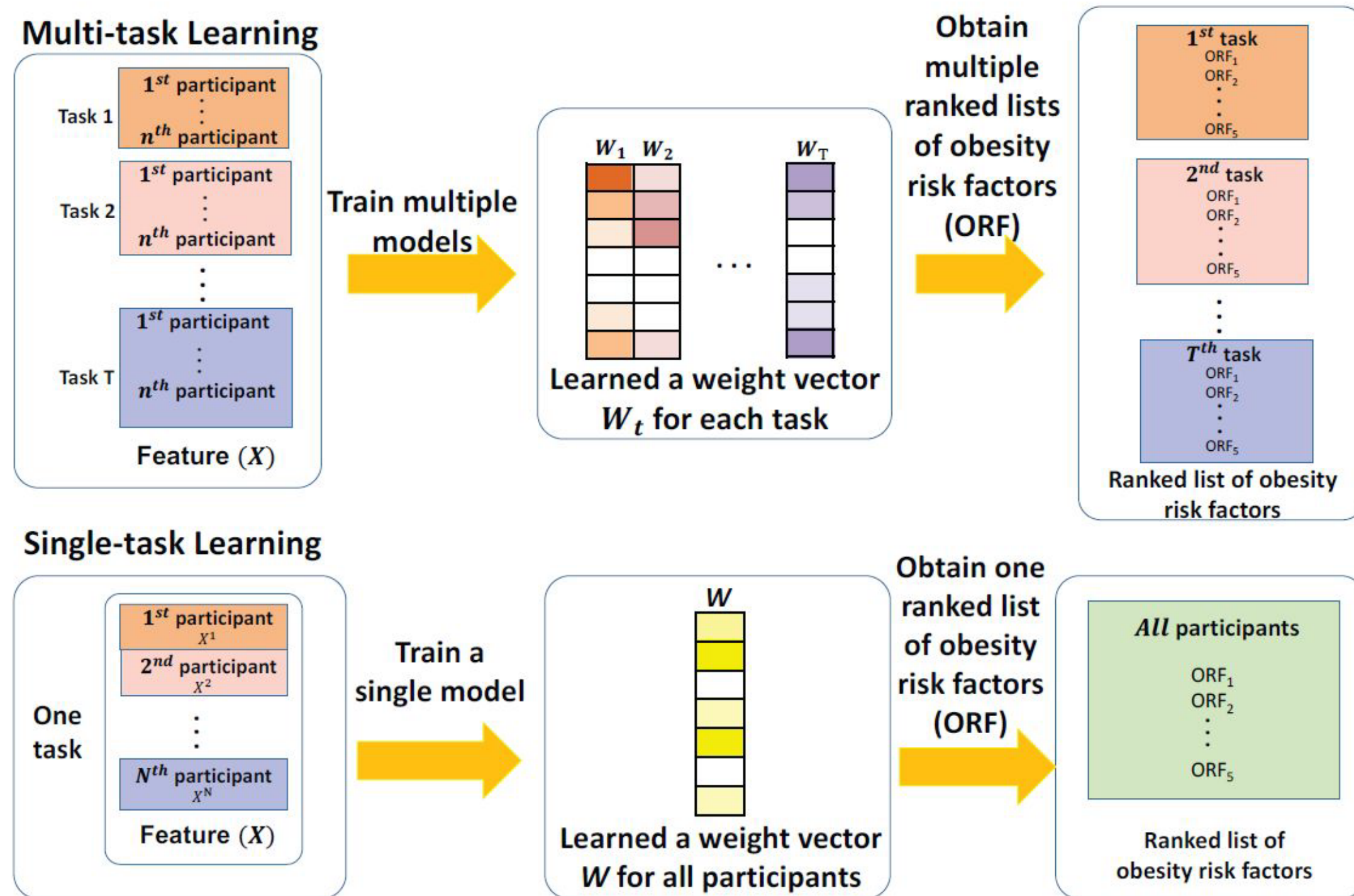


Figure 1. MTL vs STL

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 $l_{2,1}$ -Norm Regularization with Least Squares Loss (Least L21) function
- Solves the $l_{2,1}$ -norm (and the squared $l_{2,1}$ -norm) regularized multi-task least squares problem:

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

Figure 2. Top 5 Selected Risk Factors for MTL1

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

	<code>_MENT14D</code>	<code>_RFBING5</code>	<code>_STATE</code>
<code>_MENT14D</code>	1.000000	0.014568	0.000445
<code>_RFBING5</code>	0.014568	1.000000	0.009416
<code>_STATE</code>	0.000445	0.009416	1.000000

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

We would like to thank Texas State University and National Science Foundation. This material is based upon work supported by the National Science Foundation under Grant No. CNS-2149950. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Lannoy, S., Duka, T., Carbia, C., Billieux, J., Fontesse, S., Dormal, V., ... & Mauraige, P. (2021). Emotional processes in binge drinking: A systematic review and perspective. *Clinical Psychology Review*, 84, 101971.
- 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>
- 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
- 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.