

Modeling Immunogenic Neoantigens Using a Bayesian Multiple Instance Regression Model

춘계 한국통계학회

Presenter : Seongoh Park

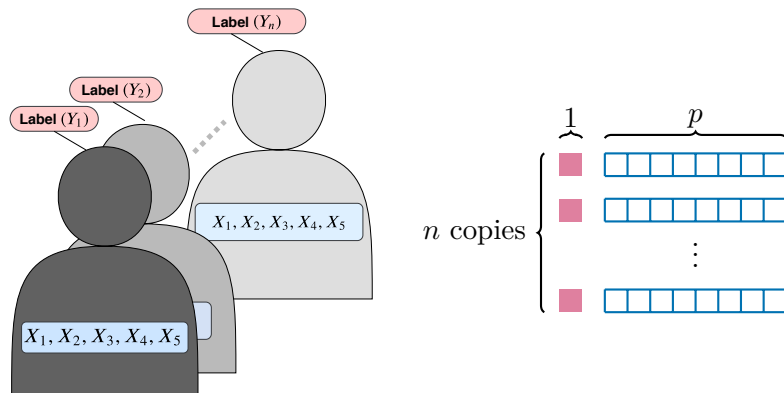
Department of Statistics, Sungshin Women's University

May 28, 2021

Section 1

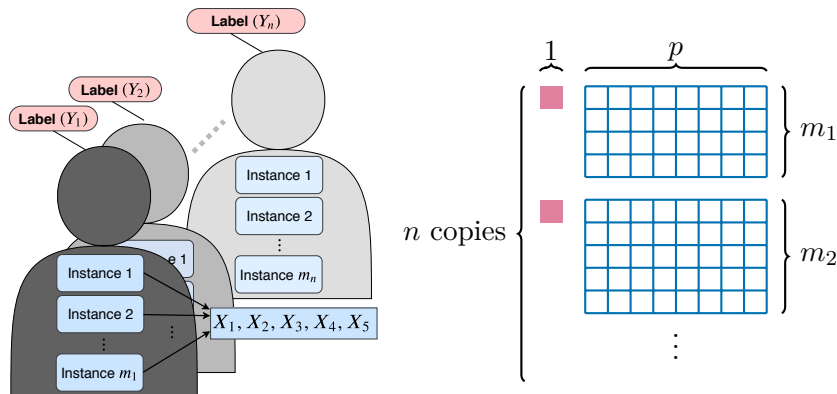
Introduction

Classical supervised learning



- ▶ One sample has one response variable (Y_i) and one corresponding feature (\mathbf{x}_i) of length fixed (p -dimensional vector)

Multiple instance learning



- ▶ One sample has **one response variable** (Y_i) and a **bag** ($\{\mathbf{x}_{ij}\}_{j=1}^{m_i}$) containing multiple feature vectors (of length p)
- ▶ Weakly supervised learning
- ▶ Each sample may have different number of instances.

Applications in MIL

- Biology and chemistry

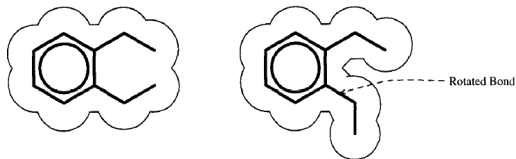


Fig. 2. The shape of a molecule changes as it rotates internal bonds. (Thin lines indicate molecular surface.)

Figure: T. G. Dietterich et al. (1997). Solving the Multiple Instance Problem with Axis-parallel Rectangles, Artificial Intelligence.

- The first MIL paper
- A molecule (a bag) can have different conformations (instances)

Applications in MIL

- Content based image retrieval

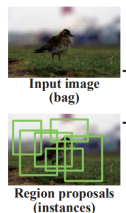


Figure: Wan, Fang et al. (2019). C-MIL: Continuation Multiple Instance Learning for Weakly Supervised Object Detection, CVPR.

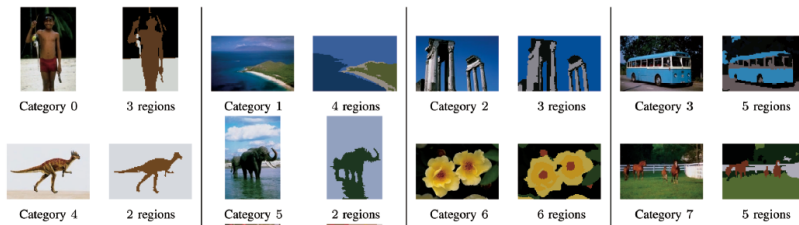


Figure: Yixin Chen et al. (2006). MILES: Multiple-Instance Learning via Embedded Instance Selection, TPAMI.

Applications in MIL

- Object localization and segmentation

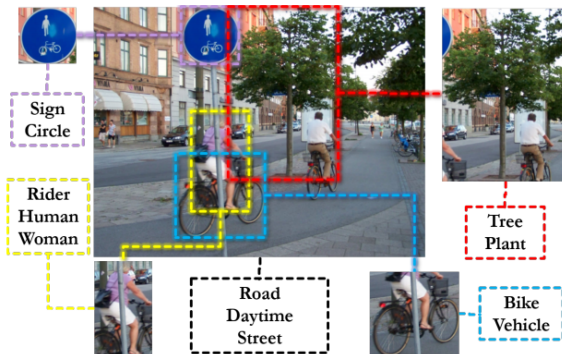


Figure: J. Wu et al. (2015). Deep multiple instance learning for image classification and auto-annotation, CVPR.

Applications in MIL

- Computer aided diagnosis and detection

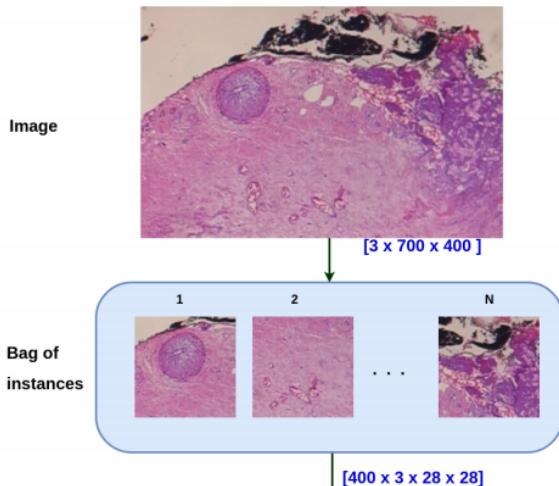


Figure: A. Patil et al. (2019). Breast Cancer Histopathology Image Classification and Localization using Multiple Instance Learning, WIECON-ECE.

Applications in MIL

► Document classification

Class	Top comment per talk (according to weights ψ_i)
inspiring	"It seems to me that the idea worth spreading of this TED Talk is inspiring and key for a full life. 'No-one else is the authority on your potential. You're the only person that decides how far you go and what you're capable of.' It seems to me that teens actually think that. As a child one is all knowing and all capable. How did we get to the (...)"
beautiful	"The beauty of the nature. It would be more interesting just integrates his thought and idea into a mobile device, like a mobile, so we can just turn on the nature gallery in any time. The paintings don't look incidental but genuinely thought out, random perhaps, but with a clear grand design behind the randomness. Drawing is an art where it doesn't (...)"
funny	"Funny story, but not as funny as a good 'knock, knock' joke. My favorite knock-knock joke of all time is Cheech & Chong's 'Dave's Not Here' gag from the early 1970s. I'm still waiting for someone to top it after all these years. [Knock, knock] 'Who is it?' the voice of an obviously stoned male answers from the other side of a door, (...)"
courageous	"I was a soldier in Iraq and part of the unit represented in this documentary. I would question anyone that told you we went over there to kill Iraqi people. I spent the better part of my time in Iraq protecting the Iraqi people from insurgents who came from countries outside of Iraq to kill Iraqi people. We protected families men, women, and (...)"

Figure: Pappas, Nikolaos et al. (2014). Explaining the Stars: Weighted Multiple-Instance Learning for Aspect-Based Sentiment Analysis, EMNLP.

Motivating data

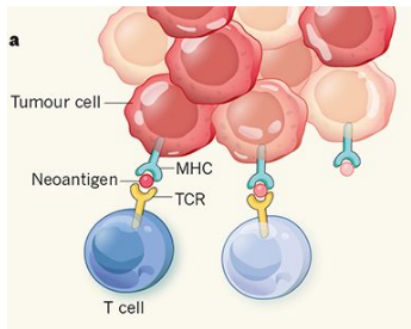
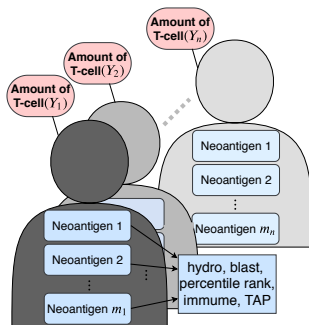
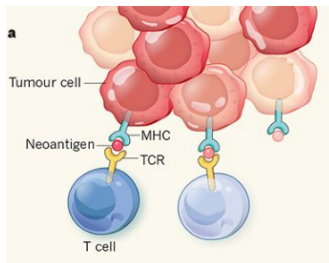


Figure: Figure 1 (a) from <https://www.nature.com/articles/d41586-017-07267-9>

- Biological question: how much T cells do neoantigens activate in a tumor cell?

Motivating data (cont'd)



- ▶ Bag: Tumor cell
- ▶ Instance: neoantigen
- ▶ Label: T cell infiltration (real-valued)

Assumptions in MIL

1. Relationship of instances with labels

- ▶ Major, necessary assumption in MIL
- ▶ Will be discussed soon

2. Prediction level

- ▶ Bag-level (majority of cases)
 - ▶ Perform prediction at bag-level
 - ▶ Do not care about the instance-level prediction
- ▶ Instance-level
 - ▶ Predict individual instances and use them for bag-level labels
 - ▶ The goal may not match with the former

Label Assumptions in MIC

1. The standard MIL assumption

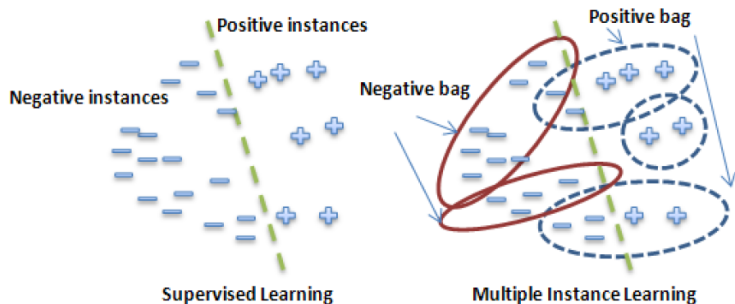


Figure: Kumar, Jayant et al. (2011). Document Image Classification and Labeling Using Multiple Instance Learning. ICDAR.

- Enough to identify at least one positive instance in each positive bag

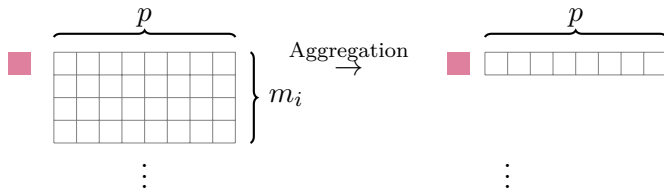
Label Assumptions in MIC

2. Relaxation of the standard MIL assumption

- ▶ Positive bags are identified by the distribution of multiple (positive) instances
- ▶ E.g. Positive bag \Leftrightarrow witness rate $> \alpha$ (e.g. traffic jam detection)

Label Assumptions in MIR

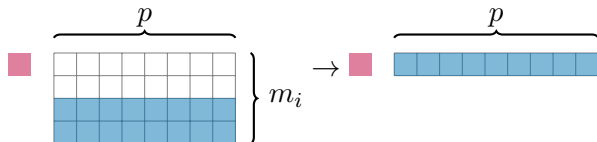
1. The metadata-based flavor



- ▶ A classical regression model can be applied to the “meta” instance
- ▶ Cannot model individual instances
- ▶ Cannot capture their relationship with bag-level responses

Label Assumptions in MIR

2. The instance-based flavor



- ▶ In each bag, there is one or more “primary” instances
- ▶ Primary instances determine the bag-level response
- ▶ Regularization techniques (e.g. ℓ_1 , ℓ_2 -penalty), sparse weights on instances
- ▶ Assign a latent binary variable to each instance where 1 corresponds to a primary instance and 0 otherwise

Related work in MIL

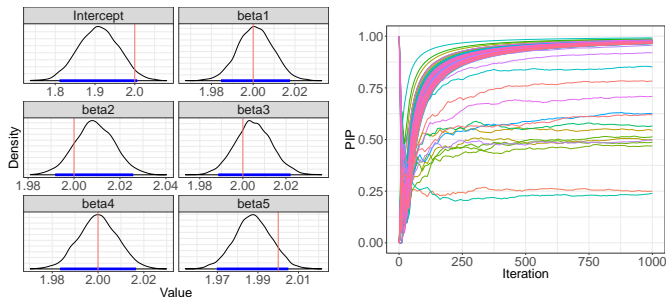
- ▶ MIC: a plethora of MIL algorithms from computer science
 - ▶ Seminal work [Dietterich et al., 1997]
 - ▶ Surveys [Zhou, 2004, Babenko, 2008, Foulds and Frank, 2010, Amores, 2013, Carbonneau et al., 2018]
 - ▶ A recent book [Herrera et al., 2016]
 - ▶ Deep neural networks
[Sun et al., 2016, Ilse et al., 2018, Ramsauer et al., 2021]

Related work in MIL

- ▶ MIR has relatively scarce literature
 1. Until now, proposed are numerical optimization methods, or even simpler algorithm-driven techniques
 - ▶ Least square estimation with a ridge penalty
[Davis et al., 2007, Lu et al., 2011]
 - ▶ Support vector regression [EL-Manzalawy et al., 2011]
 - ▶ No subsequent statistical inference
 2. Moreover, the main focus is to improve predictive performance (e.g. [Du and Zare, 2019])
 - ▶ No consideration on selection of the primary instances

Contribution of this work

- Bayesian modeling \Rightarrow Inference based on posterior distribution
- Primary instance selection (instance-based flavor)



Section 2

Bayesian Multiple Instance Regression

Data and modeling

$$Y_i = \beta_0 + \beta^T \left(\sum_{j=1}^{m_i} \delta_{ij} \mathbf{x}_{ij} \right) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2),$$

- ▶ i for a bag,
 $j = j(i)$ for an instance in bag i
- ▶ $\mathbf{x}_{ij} \in \mathbb{R}^p$: a covariate vector of instance j in bag i
- ▶ δ_{ij} : a binary indicator (0:non-primary, 1:primary)
- ▶ $\sum_{j=1}^{m_i} \delta_{ij} = 1$ (a single primary instance assumption used in [Ray and Page, 2001]) \Rightarrow immunodominance!

Related work - linear regression model

$$Y_i = \beta_0 + \beta^T \left(\sum_{j=1}^{m_i} \delta_{ij} \mathbf{x}_{ij} \right) + \epsilon_i$$

- ▶ δ_{ij} : unobservable, latent variables
 - ▶ EM algorithm: [Wang et al., 2012] \Rightarrow lack of inference
 - ▶ Bayesian modeling: [Subramanian et al., 2016] (the largest cluster in a bag explains Y) \Rightarrow not realistic in many applications
- ▶ Consider continuous weights w_{ij} instead of δ_{ij}
 - ▶ $w_{ij} \in [0, 1]$ and $\sum_j w_{ij} = 1$
 - ▶ [Pappas and Popescu-Belis, 2014, Lu et al., 2011, Wagstaff and Lane, 2007] \Rightarrow lack of inference

Prior specification - coefficients, variances

$$Y_i = \beta_0 + \boldsymbol{\beta}^T \left(\sum_{j=1}^{m_i} \delta_{ij} \mathbf{x}_{ij} \right) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2),$$

- ▶ $\beta_0 | \mu_0, \sigma_\epsilon^2, g_0 \sim N(\mu_0, \sigma_\epsilon^2 / g_0),$
- ▶ $\boldsymbol{\beta} | \boldsymbol{\mu}, \sigma_\epsilon^2, g, \boldsymbol{\Sigma} \sim N_p\left(\boldsymbol{\mu}, \frac{\sigma_\epsilon^2 \boldsymbol{\Sigma}}{g}\right),$
(the g-prior, [Zellner, 1983, Zellner, 1986]).
 - ▶ $g_0 = g = 1/n$
(the unit information prior [Kass and Wasserman, 1995])
 - ▶ $\mu_0 = \sum_{i=1}^n Y_i / n, \boldsymbol{\mu} = \mathbf{0}_p$
 - ▶ $\boldsymbol{\Sigma} = \mathbf{I}_p$
- ▶ $\sigma_\epsilon^2 | a, b \sim \text{IG}(a, b)$
 - ▶ $a = b = 0.1$ (vague)

Prior specification - binary indicators

$$Y_i = \beta_0 + \beta^T \left(\sum_{j=1}^{m_i} \delta_{ij} \mathbf{x}_{ij} \right) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2),$$

- ▶ $\boldsymbol{\delta}_i | \boldsymbol{\pi}_i \sim \text{MN}(\mathbf{1}, \boldsymbol{\pi}_i)$, where $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{im_i})^T$
(primary instance assumption, i.e. $\sum_{j=1}^{m_i} \delta_{ij} = 1$)
 - ▶ $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{im_i})^T \sim \text{Dir}(1, 1, \dots, 1)$
(i.e. $P(\boldsymbol{\delta}_i = (1, 0, \dots, 0)) = \dots = P(\boldsymbol{\delta}_i = (0, \dots, 0, 1)) = 1/m_i$)

Gibbs sampling

1. $\beta_0 | \dots \sim N \left(\frac{1}{n} \sum_{i=1}^n \left(Y_i - \boldsymbol{\beta}^T \mathbf{x}_i^\delta \right) \cdot \frac{n}{n + g_0} + \mu_0 \cdot \frac{g_0}{n + g_0}, \quad \frac{\sigma_\epsilon^2}{n + g_0} \right)$ and

$\boldsymbol{\beta} | \dots \sim N_p \left((\mathbf{A} + \mathbf{B})^{-1} \mathbf{A} \boldsymbol{\gamma} + (\mathbf{A} + \mathbf{B})^{-1} \mathbf{B} \boldsymbol{\mu}, \quad (\mathbf{A} + \mathbf{B})^{-1} \right)$ where
 $\mathbf{A} = (\mathbf{X}^\delta)^T \mathbf{X}^\delta$, $\mathbf{B} = g \boldsymbol{\Sigma}^{-1}$, and $\boldsymbol{\gamma} = ((\mathbf{X}^\delta)^T \mathbf{X}^\delta)^{-1} (\mathbf{X}^\delta)^T (\mathbf{Y} - \beta_0 \mathbf{1}_n)$

2. $\sigma_\epsilon^2 | \dots \sim \text{IG} \left(a + \frac{n + p + 1}{2}, \quad b + \frac{1}{2} \sum_{i=1}^n (Y_i - \beta_0 - \boldsymbol{\beta}^T \mathbf{x}_i^\delta)^2 + \frac{g}{2} (\boldsymbol{\beta} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) + \frac{g_0}{2} (\beta_0 - \mu_0)^2 \right)$

3. $\boldsymbol{\delta}_i | \dots \sim \text{MN} \left(1, \tilde{\boldsymbol{\pi}}_i \right)$ with the probability vector

$\tilde{\boldsymbol{\pi}}_i = (\tilde{\pi}_{i1}, \dots, \tilde{\pi}_{im_i})^T$ given by

$$\tilde{\pi}_{ij} \propto \pi_{ij} \exp \left\{ - (Y_i - \beta_0 - \boldsymbol{\beta}^T \mathbf{x}_{ij})^2 / 2\sigma_\epsilon^2 \right\}, \quad j = 1, \dots, m_i,$$

Inference in training samples

- ▶ In training samples (or bags), posterior samples (say B many) are used for parameter estimation;

$$\hat{\beta}_0 = \frac{\sum_{b=1}^B \beta_0^{(b)}}{B}, \quad \hat{\beta} = \frac{\sum_{b=1}^B \beta^{(b)}}{B}, \quad \hat{\pi}_{ij} = \frac{\sum_{b=1}^B \delta_{ij}^{(b)}}{B}, \forall i, j$$

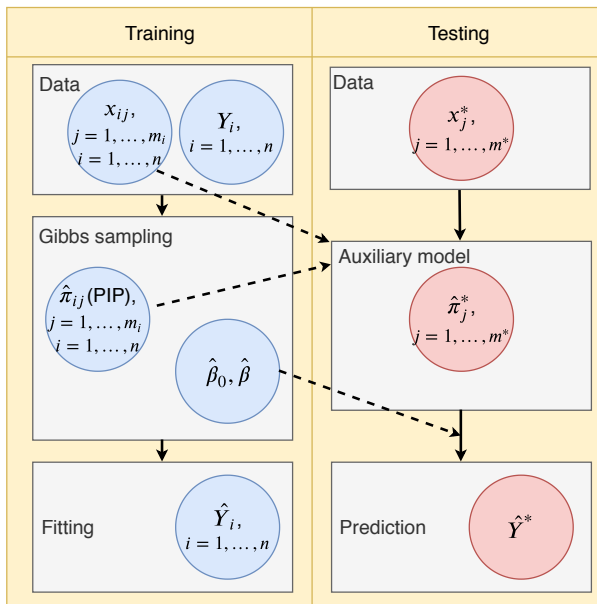
- ▶ The set of (estimated) primary instances in bag i ,

$$J_i^{max} = \{\operatorname{argmax}_{j \in \{1, \dots, m_i\}} \hat{\pi}_{ij}\}$$

- ▶ Then, the fitted value of Y_i is

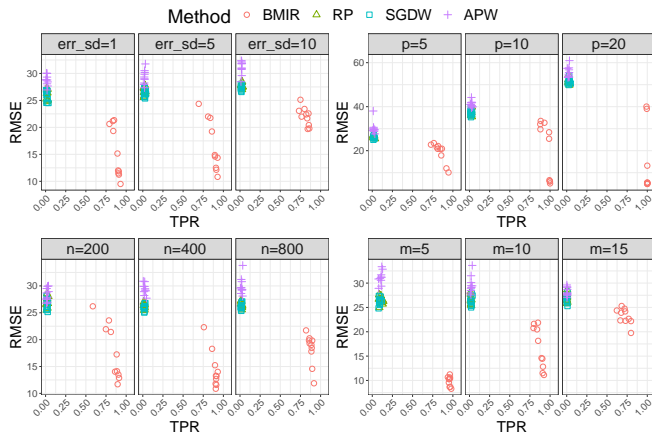
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}^T \sum_{j \in J_i^{max}} \frac{\hat{\pi}_{ij}}{\sum_{k \in J_i^{max}} \hat{\pi}_{ik}} \mathbf{x}_{ij}$$

Workflow of BMIR



Simulation study

- True positive rate (instance selection) vs. Root mean squared error (bag-level prediction)



- Comparative methods¹: proposed (BMIR), [Ray and Page, 2001] (RP),

[Pappas and Popescu-Belis, 2014] (APW), [Pappas and Popescu-Belis, 2017] (SGDW)

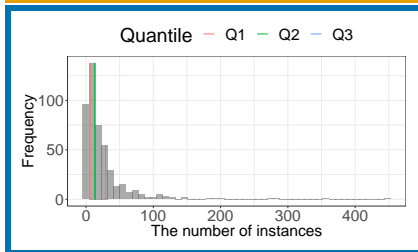
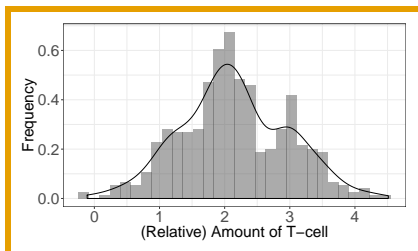
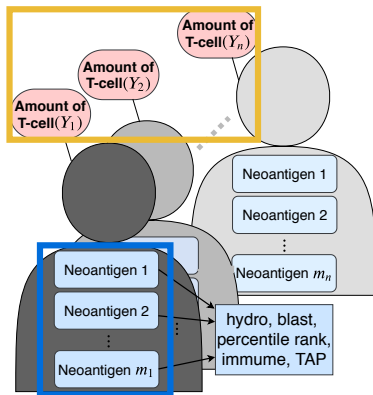
¹Unless codes are publicly available, such methods are not included in the numerical study 23 / 28

Section 3

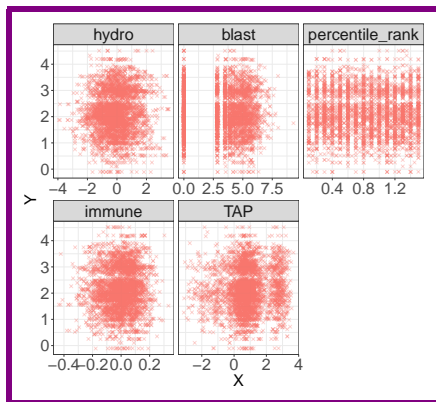
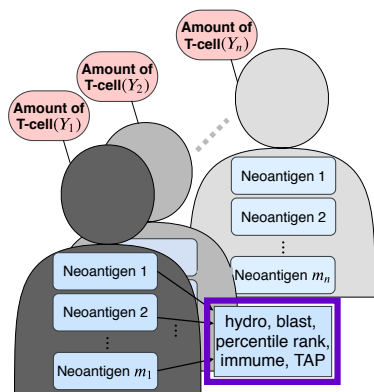
Application - immunogenicity of neoantigens

Data description

- $n = 435$ samples (kidney cancer) from four different cohorts



Data description - covariates of instances



- ▶ hydro : hydrophobicity
- ▶ blast : similarity to pathogenic epitopes
- ▶ percentile rank : rank of binding affinities to MHC molecules
- ▶ immune : an immunogenicity score previously established for class I neoantigens
- ▶ TAP : transport efficiency of a neoantigen

Validation results

- ▶ Posterior mean $\beta = (0.035, 0.021, -0.018, 0.721, 0.032, -0.093)^T$:
the directions are consistent with expectations from biological insights.
- ▶ (The instance level) $n = 65$ neoantigens (12 are immunogenic and 53 are non-immunogenic)

Method	BMIR	RP	APW	SGDW	MICIR
ROC	0.697	0.237	0.270	0.566	0.381
Precision-recall	0.330	0.116	0.120	0.199	0.142

- ▶ (The bag level) $n = 246$ in *in vivo* tumor samples (bags) from LUAD patient cohort
 - ▶ Three groups of LUAD patients based on T cell infiltration levels (high/middle/low)
 - ▶ $\hat{Y}^* \text{ (low)} < \hat{Y}^* \text{ (middle)} < \hat{Y}^* \text{ (high)}$ (p-value=0.075)

Conclusion

- ▶ We developed a Bayesian regression model under the multiple instance learning framework based on primary instance assumption
- ▶ Our model can select the primary instance(s) and predict the bag-level responses
- ▶ Quantative modeling for immunogenicity of neoantigens
- ▶ Our work is a good place to start developing statistical models for the multiple instance regression
- ▶ Published at Statistical Methods in Medical Research in 2020
R code: <https://github.com/inmybrain/BayesianMIR>

Future work

- ▶ Bayesian model for classification (with D. Xiong, X. Wang (SMU))

- ▶ A probit regression model for a binary Y

$$P(Y = 1 | \{\mathbf{x}_j\}_{j=1}^m, \{\delta_j\}_{j=1}^m, \boldsymbol{\beta}) = \Phi \left(\beta_0 + \boldsymbol{\beta}^T \left(\sum_{j=1}^{m_i} \delta_{ij} \mathbf{x}_{ij} \right) \right)$$

- ▶ Data augmentation ensures the Gibbs sampling for $\boldsymbol{\beta}$
 - ▶ Simultaneous selection procedure for both variables and instances (with J. Kim (Yonsei U.), J. Lim (SNU))
 - ▶ $\beta_k | \gamma_k \sim \gamma_k N(0, \sigma_\beta^2 \tau_{in}^2) + (1 - \gamma_k) N(0, \sigma_\beta^2 \tau_{out}^2), k = 1, \dots, p$
[George and McCulloch, 1993]
 - ▶ Shotgun Stochastic Search: effectively search a neighborhood in $\{0, 1\}^p$ -space [Hans et al., 2007]

References

- [Amores, 2013] Amores, J. (2013).
Multiple instance classification: Review, taxonomy and comparative study.
[Artificial Intelligence](#), 201:81–105.
- [Babenko, 2008] Babenko, B. (2008).
Multiple instance learning: algorithms and applications.
[View Article PubMed/NCBI Google Scholar](#), pages 1–19.
- [Carbonneau et al., 2018] Carbonneau, M.-A., Cheplygina, V., Granger, E., and Gagnon, G. (2018).
Multiple instance learning: A survey of problem characteristics and applications.
[Pattern Recognition](#), 77:329 – 353.
- [Davis et al., 2007] Davis, J., Costa, V. S., Ray, S., and Page, D. (2007).
An integrated approach to feature invention and model construction for drug activity prediction.
In [Proceedings of the 24th International Conference on Machine Learning](#), ICML '07, pages 217–224, New York, NY, USA. ACM.
- [Dietterich et al., 1997] Dietterich, T. G., Lathrop, R. H., and Lozano-Pérez, T. (1997).
Solving the multiple instance problem with axis-parallel rectangles.
[Artificial Intelligence](#), 89(1):31–71.
- [Du and Zare, 2019] Du, X. and Zare, A. (2019).
Multiple instance choquet integral classifier fusion and regression for remote sensing applications.
[IEEE Transactions on Geoscience and Remote Sensing](#), 57(5):2741–2753.