NLVAE: A New Machine Learning Approach for Extracting and Identifying Sales-Driving Product Attributes

Zijing “Jimmy” Hu and Venkatesh Shankar

Mays Business School, Texas A&M University

March 13, 2024

Motivation

- Product development

Motivation

- Product development⟶ marketing research

Motivation

- Product development ⟶ marketing research
    - Traditional methods are expensive and time-consuming

Motivation

- Product development⟶ marketing research
  - Traditional methods are expensive and time-consuming
  - Can firms use publicly available data (e.g., ratings/reviews)?

Motivation

- Product development$\longrightarrow$ marketing research
  - Traditional methods are expensive and time-consuming
  - Can firms use publicly available data (e.g., ratings/reviews)?
    - Rich information

Motivation

- Product development $\longrightarrow$ marketing research
  - Traditional methods are expensive and time-consuming
  - Can firms use publicly available data (e.g., ratings/reviews)?
    - Rich information
    - Continuously updating

Motivation

- Product development ⟶ marketing research
  - Traditional methods are expensive and time-consuming
  - Can firms use publicly available data (e.g., ratings/reviews)?
    - Rich information
    - Continuously updating
    - Cheap
    - ......

Motivation

- Product development ⟶ marketing research
  - Traditional methods are expensive and time-consuming
  - Can firms use publicly available data (e.g., ratings/reviews)?
    - Rich information
    - Continuously updating
    - Cheap
    - ......
    - **Bias**

Motivation

- Product development$\longrightarrow$ marketing research
  - Traditional methods are expensive and time-consuming
  - Can firms use publicly available data (e.g., ratings/reviews)?
    - Rich information
    - Continuously updating
    - Cheap
    - ......
    - **Bias**
  - We need advanced techniques to better utilize this data and facilitate effective and efficient marketing research.

Gaps in Relevant Literature

- Marketing research by leveraging publicly available structured and unstructured data (Lee and Bradlow 2011; Tirunillai and Tellis 2014; Timoshenko and Hauser 2019; Toubia et al. 2019; Dhillon and Aral 2021; Chakraborty et al. 2022; Zhang and Luo 2023)

Gaps in Relevant Literature

- Marketing research by leveraging publicly available structured and unstructured data (Lee and Bradlow 2011; Tirunillai and Tellis 2014; Timoshenko and Hauser 2019; Toubia et al. 2019; Dhillon and Aral 2021; Chakraborty et al. 2022; Zhang and Luo 2023)
- Biases in public reputation systems (Li and Hitt 2008; Ghose et al. 2012; Nosko and Tadelis 2015; Dai et al. 2018; He et al. 2022)

Gaps in Relevant Literature

- Marketing research by leveraging publicly available structured and unstructured data (Lee and Bradlow 2011; Tirunillai and Tellis 2014; Timoshenko and Hauser 2019; Toubia et al. 2019; Dhillon and Aral 2021; Chakraborty et al. 2022; Zhang and Luo 2023)
- Biases in public reputation systems (Li and Hitt 2008; Ghose et al. 2012; Nosko and Tadelis 2015; Dai et al. 2018; He et al. 2022)
- Limited research studies bias correction when utilizing structured and unstructured data from reputation systems.
  $\longrightarrow$ **This study**

Research Questions

- How can we extract and **accurately measure** product attributes from structured and unstructured data on consumer evaluation of products?
- Which extracted product attributes contribute most to sales?

Our Contributions

- To the literature on new product development
  - A general model to accurately measure important yet unobservable attributes from publicly available data.
  - Some attributes that contribute to better customer feedback might not necessarily drive sales.

Our Contributions

- To the literature on new product development
  - A general model to accurately measure important yet unobservable attributes from publicly available data.
  - Some attributes that contribute to better customer feedback might not necessarily drive sales.
- To the literature on ML applications in marketing
  - A theory-driven deep learning architecture that overcomes bias and enhances explainability.

Recovering the True Value with Biased Measures

We start with Theorem (Hu and Schennach 2008)

Recovering the True Value with Biased Measures

We start with Theorem (Hu and Schennach 2008)

If we can find three measures of a variable of interest that are correlated through only the true value (**CI condition**), we can uniquely identify the distribution of the variable.

Recovering the True Value with Biased Measures

We start with Theorem (Hu and Schennach 2008)

If we can find three measures of a variable of interest that are correlated through only the true value (**CI condition**), we can uniquely identify the distribution of the variable.

$$
\begin{aligned}
&f_{X_1 X_2 X_3}(x_1, x_2, x_3) \\
&\qquad = \int_{\mathcal{Z}} f_{X_1|Z}\left(x_1 \mid z\right) f_{X_2|Z}\left(x_2 \mid z\right) f_{X_3 Z}\left(x_3, z\right) dz
\end{aligned}
$$

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

$$f_{X_1,X_2,...,X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

$$f_{X_1,X_2,...,X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

- Limitations: Identification and convergence

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

$$f_{X_1,X_2,...,X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

- Limitations: Identification and convergence
- We propose **N**on-**L**atent **V**ariational **A**uto**E**ncoder (NLVAE)

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

$$f_{X_1,X_2,...,X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

- Limitations: Identification and convergence
- We propose **N**on-**L**atent **V**ariational **A**uto**E**ncoder (NLVAE)
  - Find a mapping from observable to unobservable (non-latent)

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

$$f_{X_1,X_2,...,X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

- Limitations: Identification and convergence
- We propose **N**on-**L**atent **V**ariational **A**uto**E**ncoder (NLVAE)
  - Find a mapping from observable to unobservable (non-latent)
  - Complete data generation process
    measure $\xrightarrow{encoder}$ unobservable true value $\xrightarrow{decoder}$ measure

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

$$f_{X_1,X_2,...,X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

- Limitations: Identification and convergence
- We propose **N**on-**L**atent **V**ariational **A**uto**E**ncoder (NLVAE)
  - Find a mapping from observable to unobservable (non-latent)
  - Complete data generation process
    measure $\xrightarrow{encoder}$ unobservable true value $\xrightarrow{decoder}$ measure
  - CI condition as a regularization for the encoding space

Recovering the True Value Using Neural Network (NN)

- Hu et al. (2023): Predict $\hat{Z} = \mathrm{NN}(X_1, X_2, ..., X_N)$ that satisfy the CI condition:

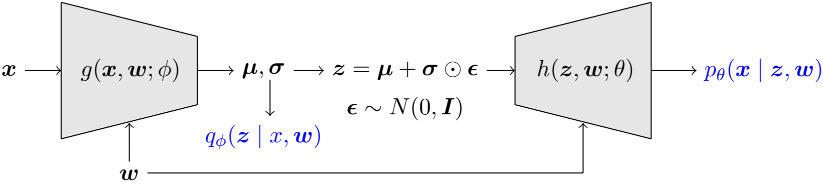
$$f_{X_1, X_2, ..., X_N|\hat{Z}} = f_{X_1|\hat{Z}} f_{X_1|\hat{Z}} \cdots f_{X_N|\hat{Z}}$$

- Limitations: Identification and convergence
- We propose **N**on-**L**atent **V**ariational **A**uto**E**ncoder (NLVAE)
  - Find a mapping from observable to unobservable (non-latent)
  - Complete data generation process
    measure $\xrightarrow{encoder}$ unobservable true value $\xrightarrow{decoder}$ measure
  - CI condition as a regularization for the encoding space
  - Mild assumptions to ensure identification

Conditional Variational Autoencoder (CVAE)

We use the architecture of CVAE (Sohn et al. 2015)

- CVAE reconstructs only one input that contains most information
- Other inputs are treated as conditions in encoding and decoding


$\boldsymbol{x}$
$g(\boldsymbol{x}, \boldsymbol{w}; \phi)$
$\boldsymbol{\mu}, \boldsymbol{\sigma}$
$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$
$\boldsymbol{\epsilon} \sim N(0, \boldsymbol{I})$
$h(\boldsymbol{z}, \boldsymbol{w}; \theta)$
$p_\theta(\boldsymbol{x} \mid \boldsymbol{z}, \boldsymbol{w})$
$q_\phi(\boldsymbol{z} \mid x, \boldsymbol{w})$
$\boldsymbol{w}$

The optimization target is given by the (inverted) ELBO:

$$\mathcal{F}_{\mathrm{CVAE}}(\theta, \phi) = \frac{1}{N} \sum_{i=1}^{N} \left[-\log p_\theta \left(x_i \mid \boldsymbol{z}_i, \boldsymbol{w}_i \right) + \mathrm{D}_{\mathrm{KL}} \left[q_\phi \left(\boldsymbol{z}_i \mid x_i, \boldsymbol{w}_i \right) \| \Pr \left(\boldsymbol{z}_i \right) \right] \right]$$

NLVAE = CVAE + CI Condition

The CI restriction over the enconding space leads to a new optimization target:

$$\min \mathcal{F}_{\mathrm{CVAE}}(\theta, \phi)\ s.t.\ \mathrm{D}_{\mathrm{KL}}\left[f_{X\boldsymbol{W}_j Z_j} || f_{XZ_j} \prod_{k=1}^{K} f_{W_{jk}|Z_j}\right] \leq \varepsilon$$

NLVAE = CVAE + CI Condition

The CI restriction over the enconding space leads to a new optimization target:

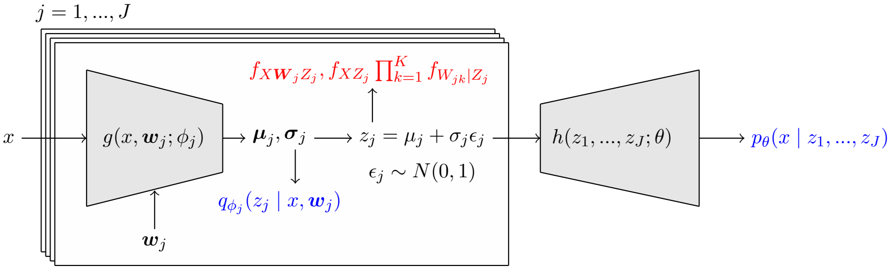
$$\min \mathcal{F}_{\mathrm{CVAE}}(\theta, \phi)\ s.t.\ \mathrm{D}_{\mathrm{KL}}\left[f_{X\boldsymbol{W}_j Z_j} \| f_{XZ_j} \prod_{k=1}^{K} f_{W_{jk}|Z_j}\right] \leq \varepsilon$$

Incorporating the Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{aligned}\mathcal{F}_{\mathrm{KKT}}(\theta, \phi; \beta) = \frac{1}{N} \sum_{i=1}^{N} \Big[&-\log p_\theta\left(x^{(i)} \mid \boldsymbol{z}^{(i)}\right) + \mathrm{D}_{\mathrm{KL}}\left[q_\phi\left(\boldsymbol{z}^{(i)} \mid x^{(i)}, \boldsymbol{w}_1^{(i)}, \ldots, \boldsymbol{w}_J^{(i)}\right) \| \operatorname{Pr}\left(\boldsymbol{z}^{(i)}\right)\right]\Big] \\ &+ \beta \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \mathrm{D}_{\mathrm{KL}}\left[f_{X\boldsymbol{W}_j Z_j}(x^{(i)}, \boldsymbol{w}_j^{(i)}, z_j^{(i)}) \| f_{XZ_j}(x^{(i)}, z_j^{(i)}) \prod_{k=1}^{K} f_{W_{jk}|Z_j}(w_{jk}^{(i)}, z_j^{(i)})\right]\end{aligned}$$

β is a hyperparameter, similar to β-VAE (Higgins et al. 2017)

Model Architecture of NLVAE



$j = 1, ..., J$

x

$g(x, \boldsymbol{w}_j; \phi_j)$

$\boldsymbol{w}_j$

$\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j$

$q_{\phi_j}(z_j \mid x, \boldsymbol{w}_j)$

$f_{X\boldsymbol{W}_j Z_j}, f_{XZ_j} \prod_{k=1}^{K} f_{W_{jk}|Z_j}$

$z_j = \mu_j + \sigma_j \epsilon_j$

$\epsilon_j \sim N(0, 1)$

$h(z_1, ..., z_J; \theta)$

$p_\theta(x \mid z_1, ..., z_J)$

Empirical context in our study:

- x: pooled measure (e.g., overall product ratings)
- $\boldsymbol{w}_j = [w_{j1}, ..., w_{jK}]$: attribute-specific measures (e.g., sentiments or mention counts of product attributes)
- z_j: true attribute score of the j-th product attribute

Simulation Settings

- Since the true values are observable, we use a hypothesized data generation process to evaluate NLVAE

Simulation Settings

- Since the true values are observable, we use a hypothesized data generation process to evaluate NLVAE
- Context and data
  - Eight product attributes
  - Pooled measure X (same for each attribute)
  - Two attribute-specific measures $(W_j, \tilde{W}_j)$
  - Each observation is given by: $(x^{(i)}, w_1^{(i)}, \tilde{w}_1^{(i)}, ..., w_8^{(i)}, \tilde{w}_8^{(i)})$

Simulation Settings

- Since the true values are observable, we use a hypothesized data generation process to evaluate NLVAE
- Context and data
  - Eight product attributes
  - Pooled measure X (same for each attribute)
  - Two attribute-specific measures $(W_j, \tilde{W}_j)$
  - Each observation is given by: $(x^{(i)}, w_1^{(i)}, \tilde{w}_1^{(i)}, ..., w_8^{(i)}, \tilde{w}_8^{(i)})$
- Goal: Decompose true attribute scores from the overall rating
  - Recover each true attribute score $z_j^{(i)}$

Data Generation Process

- The most ideal case
  - Classical, linear, and separable measurement error
  - Honest and comprehensive customer feedback + direct measures of attribute scores

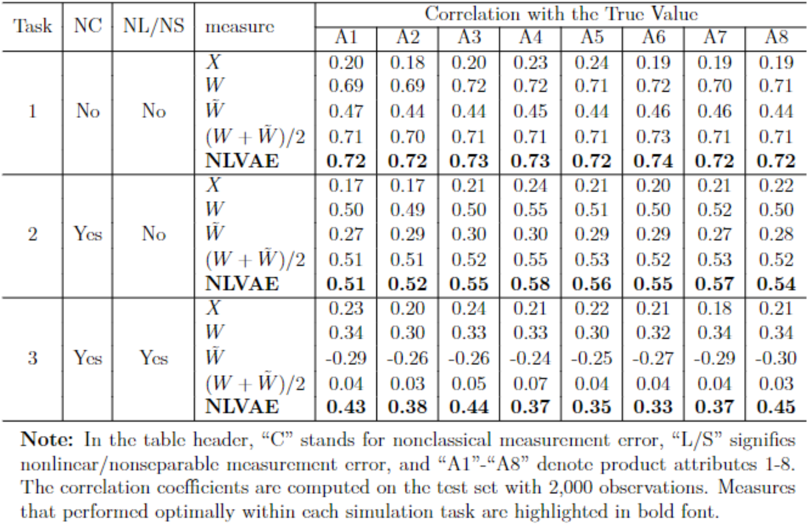
Data Generation Process

- The most ideal case
  - Classical, linear, and separable measurement error
  - Honest and comprehensive customer feedback + direct measures of attribute scores
- A worse case
  - Nonclassical, linear, and separable measurement error
  - Biased customer feedback + direct measures of attribute scores

Data Generation Process

- The most ideal case
  - Classical, linear, and separable measurement error
  - Honest and comprehensive customer feedback + direct measures of attribute scores
- A worse case
  - Nonclassical, linear, and separable measurement error
  - Biased customer feedback + direct measures of attribute scores
- The worst case (but general)
  - Nonclassical, nonlinear, and nonseparable measurement error
  - Biased customer feedback + indirect measures of attribute scores

Simulation Results



Task	NC	NL/NS	measure	Correlation with the True Value							
				A1	A2	A3	A4	A5	A6	A7	A8
1	No	No	X	0.20	0.18	0.20	0.23	0.24	0.19	0.19	0.19
			W	0.69	0.69	0.72	0.72	0.71	0.72	0.70	0.71
			$\tilde{W}$	0.47	0.44	0.44	0.45	0.44	0.46	0.46	0.44
			$(W+\tilde{W})/2$	0.71	0.70	0.71	0.71	0.71	0.73	0.71	0.71
			NLVAE	**0.72**	**0.72**	**0.73**	**0.73**	**0.72**	**0.74**	**0.72**	**0.72**
2	Yes	No	X	0.17	0.17	0.21	0.24	0.21	0.20	0.21	0.22
			W	0.50	0.49	0.50	0.55	0.51	0.50	0.52	0.50
			$\tilde{W}$	0.27	0.29	0.30	0.30	0.29	0.29	0.27	0.28
			$(W+\tilde{W})/2$	0.51	0.51	0.52	0.55	0.53	0.52	0.53	0.52
			NLVAE	**0.51**	**0.52**	**0.55**	**0.58**	**0.56**	**0.55**	**0.57**	**0.54**
3	Yes	Yes	X	0.23	0.20	0.24	0.21	0.22	0.21	0.18	0.21
			W	0.34	0.30	0.33	0.33	0.30	0.32	0.34	0.34
			$\tilde{W}$	-0.29	-0.26	-0.26	-0.24	-0.25	-0.27	-0.29	-0.30
			$(W+\tilde{W})/2$	0.04	0.03	0.05	0.07	0.04	0.04	0.04	0.03
			NLVAE	**0.43**	**0.38**	**0.44**	**0.37**	**0.35**	**0.33**	**0.37**	**0.45**

Note: In the table header, "C" stands for nonclassical measurement error, "L/S" signifies nonlinear/nonseparable measurement error, and "A1"-"A8" denote product attributes 1-8. The correlation coefficients are computed on the test set with 2,000 observations. Measures that performed optimally within each simulation task are highlighted in bold font.

Recovering True Attribute Scores of Video Games

- Data: Video game data from Steam
  - 117 video games from February 1, 2022, to January 31, 2023
  - Historical revenue rankings, active player counts, prices, and other relevant information

Recovering True Attribute Scores of Video Games

- Data: Video game data from Steam
  - 117 video games from February 1, 2022, to January 31, 2023
  - Historical revenue rankings, active player counts, prices, and other relevant information
- Three measures that can satisfy the CI condition
  1. Attribute mention frequencies in English-speaking regions
  2. Attribute mention frequencies in Chinese-speaking regions
  3. Overall video game rating in other regions

Recovering True Attribute Scores of Video Games

- Data: Video game data from Steam
  - 117 video games from February 1, 2022, to January 31, 2023
  - Historical revenue rankings, active player counts, prices, and other relevant information
- Three measures that can satisfy the CI condition
  1. Attribute mention frequencies in English-speaking regions
  2. Attribute mention frequencies in Chinese-speaking regions
  3. Overall video game rating in other regions
- We use Google Gemini to summarize attribute and generate attribute mention counts
  - Good capability for processing multilingual reviews
  - It embeds domain knowledge
  - Open source foundation model

Extracting 10 Video Game Attributes Using Google Gemini

#	Attribute	Description
1	Visuals	Refers to the game's art style, graphics, and overall visual presentation.
2	Music	Encompasses the game's soundtrack, sound effects, and overall audio experience.
3	Gameplay	Describes the core mechanics, controls, and objectives of the game.
4	Narrative	Refers to the game's story, characters, and overall plot.
5	Replayability	Indicates the game's ability to offer multiple playthroughs with fresh experiences.
6	Accessibility	Relates to the game's features that make it accessible to players with disabilities or varying skill levels.
7	Puzzles	Highlights the presence of brain-teasing puzzles or riddles that players must solve to progress.
8	Boss Fights	Refers to challenging encounters with powerful enemies, often at the end of levels or chapters.
9	Secrets	Indicates the game's hidden content, collectibles, or easter eggs that players can discover through exploration.
10	Community	Highlights the game's active player base and online communities where players can interact, share strategies, and create content.

Relating Product Attributes and Sales

$$y_{it}^{j} = \widehat{\boldsymbol{z}}_{it}^{\top}\boldsymbol{\beta}_1^{j} + \boldsymbol{\delta}_{it}^{\top}\boldsymbol{\beta}_2^{j} + \xi_i^{j} + \phi_t^{j} + \varepsilon_{it}^{j}.$$

- j=1: (inverted) revenue rankings; j=2: active player counts; j=3: overall ratings
- i: video game index; t: week index
- $\hat{\boldsymbol{z}}_{it}$: recovered attribute scores
- $\boldsymbol{\delta}_{it}$: covariates such as price, release years, and so on.
- ξ_i^{j}, ϕ_t^{j}: fixed effects
- ε_{it}^{j}: error term

Relating Product Attributes and Sales

Table: Attribute-Level Contributions

Variable	(1)	(2)	(3)
Visuals	-50.996***	-1125.917**	-0.001
Music	-81.227***	-2474.175***	0.005***
Gameplay	28.943***	2866.571***	-0.005***
Narrative	-55.042***	-1043.385**	0.010***
Replayability	49.078***	1280.636***	0.008***
Accessibility	8.310	-695.617**	0.005***
Puzzles	-53.203***	-2540.508***	0.014***
Boss Fights	-3.773	-170.185	0.002
Secrets	-16.414***	-2061.367***	0.015***
Community	-29.583***	-894.327**	0.007***
Control Variables	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
R-squared	0.328	0.313	0.359
N	5616	5616	5616

(1) Revenue Ranking (Inverted); (2) Active Player Count; (3) Customer Rating

Prediction Results

Table: Prediction RMSE of Models with Different Predictors

Attribute Score Predictors	(1)	(2)	(3)	(4)
Measure 1	1.029	0.548	1.195	0.609
Measure 2	1.039	0.529	1.206	0.596
Measure 3	1.120	0.524	1.518	0.676
Measure 1 + 2 + 3	0.866	0.525	1.010	0.593
NLVAE-predictors	0.880	0.514	1.078	0.578

(1) Attribute score predictors → inverted revenue rankings
(2) Attribute score predictors + control variables → inverted revenue rankings
(3) Attribute score predictors → active player counts
(4) Attribute score predictors + control variables → active player counts

Research Questions and Results

- How can we extract and accurately measure product attributes from structured and unstructured data on consumer evaluation of products?
  - We propose the NLVAE to extract and measure product attribute scores from online product overall ratings and reviews
  - NLVAE overcomes many types of biases in public data
  - NLVAE uses a theory-driven and interpretable model

Research Questions and Results

- How can we extract and accurately measure product attributes from structured and unstructured data on consumer evaluation of products?
  - We propose the NLVAE to extract and measure product attribute scores from online product overall ratings and reviews
  - NLVAE overcomes many types of biases in public data
  - NLVAE uses a theory-driven and interpretable model
- Which extracted product attributes contribute most to sales?
  - In the video game data we analyze, only Gameplay and Replayability are positively related to sales
  - Some attributes (Music, Narrative, Puzzles, Secrets, and Community) affect video game ratings but do not contribute to sales

Theoretical and Managerial Implications

- Theoretical implications
  - Important product attributes can be recovered
  - Product attributes have complex effects
  - Bias mitigation and explainability of ML models can be enhanced

Theoretical and Managerial Implications

- Theoretical implications
  - Important product attributes can be recovered
  - Product attributes have complex effects
  - Bias mitigation and explainability of ML models can be enhanced
- Managerial implications
  - Identifying sales-related yet potentially unobservable attributes for product development
  - Unique and valuable insights beyond customer feedback

Thank you!

zijinghu@tamu.edu

Appendix

Intuition of Hu and Schennach (2008)

Defining linear operators

$$L_{B|A}: \mathcal{G}(\mathcal{A}) \mapsto \mathcal{G}(\mathcal{B}) \text{ with } \left[L_{B|A}g\right](b) \equiv \int_{\mathcal{A}} f_{B|A}(b \mid a)g(a)da,$$

$$\Delta_{b;A}: \mathcal{G}(\mathcal{A}) \mapsto \mathcal{G}(\mathcal{A}) \text{ with } \Delta_{b;A}g \equiv f_{B|A}(b \mid \cdot)g(\cdot).$$

Then

$$L_{x_2;X_1|X_3} = L_{X_1|Z}\Delta_{x_2;Z}L_{Z|X_3}, \tag{1}$$

$$L_{Z|X_3} = L^{-1}_{X_1|Z}L_{X_1|X_3}, \tag{2}$$

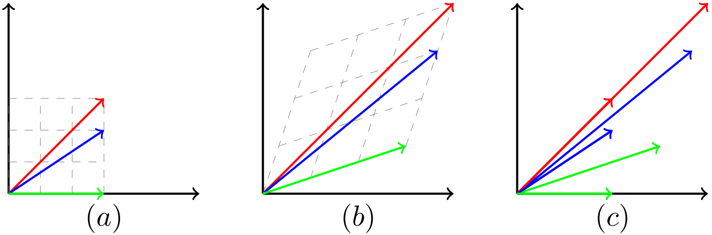
and we can use eigendecomposition to solve:

$$L_{x_2;X_1|X_3}L^{-1}_{X_1|X_3} = L_{X_1|Z}\Delta_{x_2;Z}L^{-1}_{X_1|Z},$$

Geometric Interpretation of Eigenvectors

$$L_{x_2;X_1|X_3}L^{-1}_{X_1|X_3} = L_{X_1|Z}\Delta_{x_2;Z}L^{-1}_{X_1|Z},$$

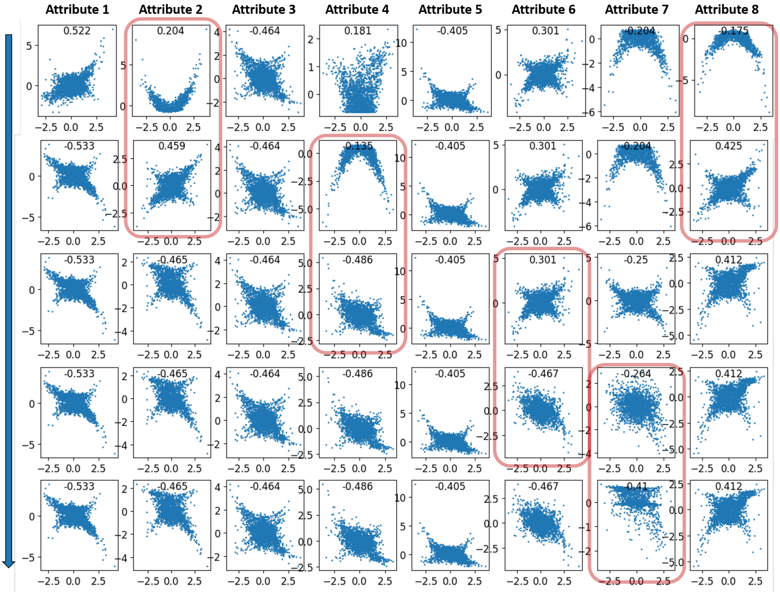
$L_{X_1|Z}$ is exactly the set of directions (vectors) that are fixed in the action (a-b) of $L_{x_2;X_1|X_3}L^{-1}_{X_1|X_3}$



Training Tricks

- The loss function of NN is non-convex. The model might get stuck in suboptimal points and generate unstable results
- The following procedure increases the robustness of our model
  - Step 0: train the whole model
  - Step 1: initialize the decoder and only one of encoders; keep other weights fixed
  - Step 2: retrain the model
    - If the new model yields a better KL divergence of the conditional independence restriction, use the new weights
    - Otherwise, use the old weights
    - Repeat Step 1-2 multiple times
  - Step 3: train the whole model until converge

Training Tricks: Visualizing the Training Process



Identification Tricks

- Challenge 1: the orders/directions of the true value
  - We impose a reasonable assumption that, given other attribute scores fixed, a better attribute score (z_j) leads to a higher overall rating (x):

$$P(\frac{\partial x}{\partial z_i} \leq 0) < P(\frac{\partial x}{\partial z_i} > 0)$$

  - Using automatic differentiation we can bootstrap the distribution of $\frac{\partial x}{\partial z_i}$ and determine the direction.
- Challenge 2: the scale of the true value
  - The regularization term $\mathrm{D_{KL}}\left[q_\phi \| \, \mathrm{p}_z\right]$ helps restrict the scale
  - Cannot fully pin it down but enough for downstream tasks

Data Generation Process (Task 1)

$$x^{(i)} = \frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{e^{\nu_{jk}^{(i)}}x_{jk}^{(i)}}{\sum_{j'=1}^{J}e^{\nu_{j'k}^{(i)}}},\ x_{jk}^{(i)} \sim \mathcal{N}(z_j^{(i)}, \sigma_x),\ \nu_{jk}^{(i)} \sim \mathcal{N}(0, \sigma_\nu),$$

$$w_j^{(i)} = \frac{1}{K}\sum_{k=1}^{K} w_{jk}^{(i)},\ w_{jk}^{(i)} \sim \mathcal{N}(z_j^{(i)}, \sigma_w),$$

$$\tilde{w}_j^{(i)} = \frac{1}{K}\sum_{k=1}^{K} \tilde{w}_{jk}^{(i)},\ \tilde{w}_{jk}^{(i)} \sim \mathcal{N}(z_j^{(i)}, \sigma_{\tilde{w}}).$$

Data Generation Process (Task 2)

$$x^{(i)} = \frac{1}{\sum_{k=1}^{K} 1_{\{\nu_{jk}^{(i)} \leq \sigma_\nu \exists j\}}} \sum_{k:\exists j\ \nu_{jk}^{(i)} \leq \sigma_\nu} \sum_{j:\nu_{jk}^{(i)} \leq \sigma_\nu} \frac{e^{\nu_{jk}^{(i)}} x_{jk}^{(i)}}{\sum_{j':\nu_{j'k}^{(i)} \leq \sigma_\nu} e^{\nu_{j'k}^{(i)}}},$$

$$w_j^{(i)} = \frac{1}{\sum_{k=1}^{K} 1_{\{w_{jk}^{(i)} \leq (\sigma_w + \sigma_z)/2\}}} \sum_{k:w_{jk}^{(i)} \leq (\sigma_w + \sigma_z)/2} w_{jk}^{(i)},$$

$$\tilde{w}_j^{(i)} = \frac{1}{\sum_{k=1}^{K} 1_{\{\tilde{w}_{jk}^{(i)} \leq (\sigma_{\tilde{w}} + \sigma_z)/2\}}} \sum_{k:\tilde{w}_{jk}^{(i)} \leq (\sigma_{\tilde{w}} + \sigma_z)/2} \tilde{w}_{jk}^{(i)}.$$

Data Generation Process (Task 3)

$$x^{(i)} = \frac{1}{\sum_{k=1}^{K} 1_{\{\nu_{jk}^{(i)} \leq \sigma_\nu \exists j\}}} \sum_{k:\exists j\ \nu_{jk}^{(i)} \leq \sigma_\nu} \sum_{j:\nu_{jk}^{(i)} \leq \sigma_\nu} \frac{e^{\nu_{jk}^{(i)}} x_{jk}^{(i)}}{\sum_{j':\nu_{j'k}^{(i)} \leq \sigma_\nu} e^{\nu_{j'k}^{(i)}}},$$

$$w_j^{(i)} = \frac{1}{\sum_{k=1}^{K} 1_{\{w_{jk}^{(i)} \leq (\sigma_w + \sigma_z)/2\}}} \sum_{k:w_{jk}^{(i)} \leq (\sigma_w + \sigma_z)/2} (w_{jk}^{(i)} + \sigma_w + \sigma_z)^2,$$

$$\tilde{w}_j^{(i)} = \frac{1}{\sum_{k=1}^{K} 1_{\{\tilde{w}_{jk}^{(i)} \leq (\sigma_{\tilde{w}} + \sigma_z)/2\}}} \sum_{k:\tilde{w}_{jk}^{(i)} \leq (\sigma_{\tilde{w}} + \sigma_z)/2} \frac{1}{\tilde{w}_{jk}^{(i)}}.$$

Frequency and Correlation of Attribute Mentions

Attributes	Freq	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
1. Visuals	0.17	1.00	0.59	0.18	0.30	0.11	0.09	0.13	0.21	0.18	0.02
2. Music	0.13	0.59	1.00	0.14	0.36	0.12	0.08	0.14	0.22	0.19	0.04
3. Gameplay	0.65	0.18	0.14	1.00	0.06	0.18	0.06	0.11	0.21	0.09	0.06
4. Narrative	0.25	0.30	0.36	0.06	1.00	0.04	0.03	0.14	0.19	0.16	-0.02
5. Replayability	0.09	0.11	0.12	0.18	0.04	1.00	0.06	0.03	0.13	0.17	0.16
6. Accessibility	0.02	0.09	0.08	0.06	0.03	0.06	1.00	0.04	0.06	0.07	0.05
7. Puzzles	0.04	0.13	0.14	0.11	0.14	0.03	0.04	1.00	0.13	0.21	0.01
8. Boss Fights	0.10	0.21	0.22	0.21	0.19	0.13	0.06	0.13	1.00	0.26	0.04
9. Secrets	0.04	0.18	0.19	0.09	0.16	0.17	0.07	0.21	0.26	1.00	0.07
10. Community	0.04	0.02	0.04	0.06	-0.02	0.16	0.05	0.01	0.04	0.07	1.00