Identifying Always-the-Same-Rating Reviewers in a One-sided-Review System Using Big Data Analytics

Jikhan Jeong, Ph.D.

Florida Polytechnic University

2022 AEA Poster Presentation

1. Introduction

Research Question

- **One-sided review systems anyone can write product reviews as a buyer without providing personal information.**
- **'Always-the-same-rating reviewers' (ASRs) might**

  → lessen the informativeness of average measures of product quality (e.g., average star ratings).

  → decrease the credibility of online product review and consumer surplus.
- **(Q1) Identifying Always-the-Same-Rating Reviewers (ASRs) in a One-sided-Review (Amazon.com).**

  → (Big Data Analytics) Calculating all reviews using HPC

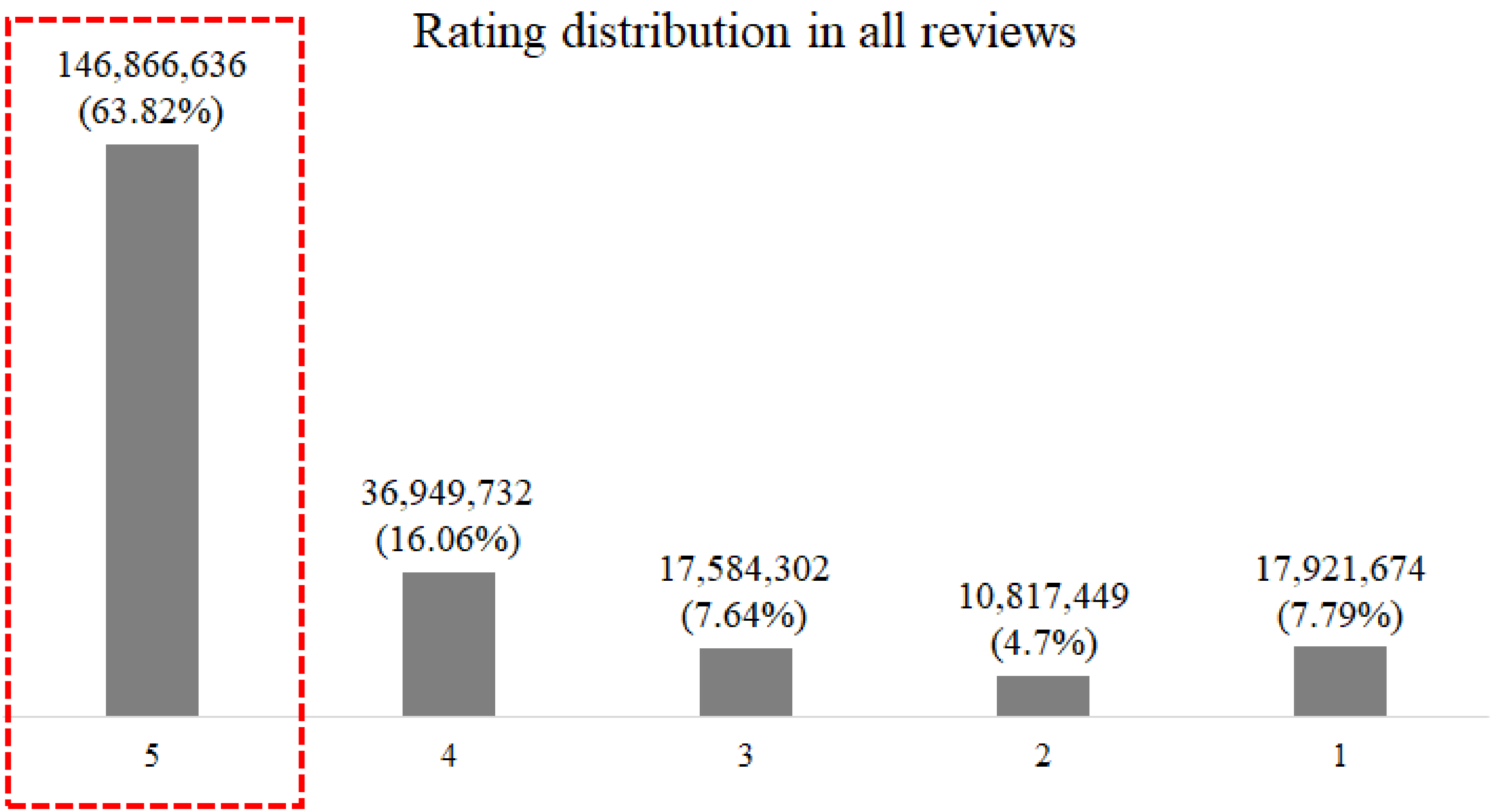
  → (AI) Classification using deep learning (i.e., NLP)
- **(Q2) Identifying the characteristics of reviews written by ASRs**

  → (Binary logit models) Identifying the determinants of purchased-verified ASRs' reviews

  → (Binary logit models) Identifying the key determinants of non-verified ASRs' reviews

Data

- The initial Amazon product review dataset contains 233M reviews for 15M products written by 102M reviewers on Amazon.com between May 1996 and Oct 2018 (Ni, Li, and McAuley 2019).



Literature Review

- **Hu and Pavlou (2009)** suggested that **two self-selection biases** increase **the chance of a positive skewed (J-shaped)** rating distribution in online product reviews.

→ The first one is **"purchasing bias",** which might exist because consumers have positive expectations about a product and have a chance to write reviews, while consumers who have a negative ex ante expectation of the product may not buy the product and do not have a chance to write a review.

→ The second self-selection bias is **"underreporting bias",** which might exist because reviewers are likely to write a review when they are either very satisfied or very unsatisfied with the reviewed products but do not bother otherwise.

- **Hu, Pavlou, and Zhang (2006)** and **Hu, Pavlou, and Zhang (2017)** studied "polarity self-selection bias" in online product reviews.

Literature Review

- **Reimers and Waldfogel (2020)** suggested that information in prior online product reviews (e.g., average star rating) can improve consumer welfare when making purchases.
- **De Langhe, Fernbach, and Lichtenstein (2016)** demonstrated that consumers rely on average star ratings to estimate product quality with and without enough prior reviews.
- **Schoenmueller, Netzer, and Stahl (2020)** found that a higher proportion of 5- and 1-star ratings (extreme ratings) lessens the informativeness of the average review measures (e.g., average star rating).
- **Karaman (2020)** also defined "extremity bias" in online reviews, stating that reviewers cannot represent the population of consumers for a reviewed product or service

→ **Polarity self-selection biases** in reviews and promotional reviews can reduce the usefulness and credibility of information contained in reviews, thereby reducing their usefulness.

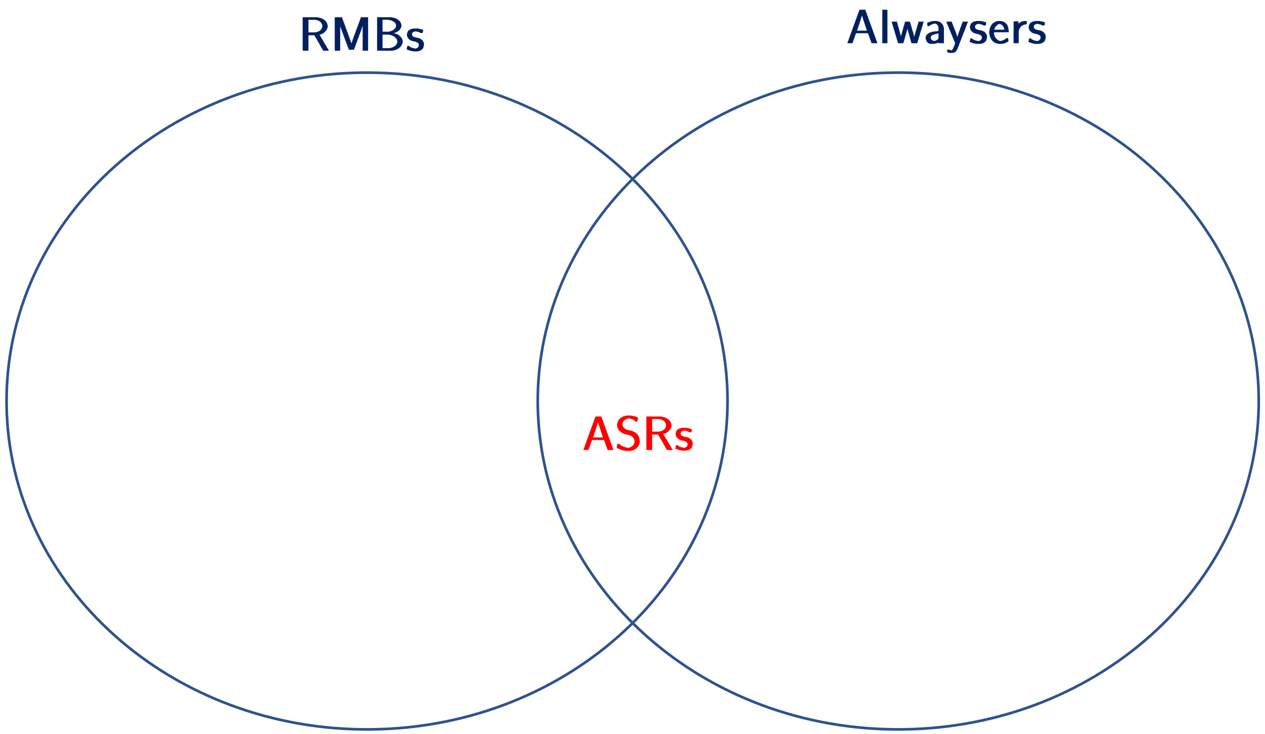
2. Big data analysis : descriptive study

Big data analysis : descriptive study

- The study assumed that ASRs **are the reviewers that write more than twelve reviews with the same star rating**.
- If the probability of the majority rating in the five-scale star-ratings is 0.7, the probability that a reviewer independently writes reviews with the same majority star rating level in thirteen consecutive reviews is 0.0097 (less than 1%).
- First, 'reviewers write reviews more than the bar (**RMBs**)' denotes reviewers that have written more than twelve reviews.
- **'Alwaysers'** denotes the reviewers give star rating at the same level for all reviewed products in the given category.
- ASRs in a category are simply the intersection between 'RMBs' and 'Alwayers'.

Big data analysis : descriptive study

- ASRs in a category are simply the intersection between 'RMBs' and 'Alwaysers'.



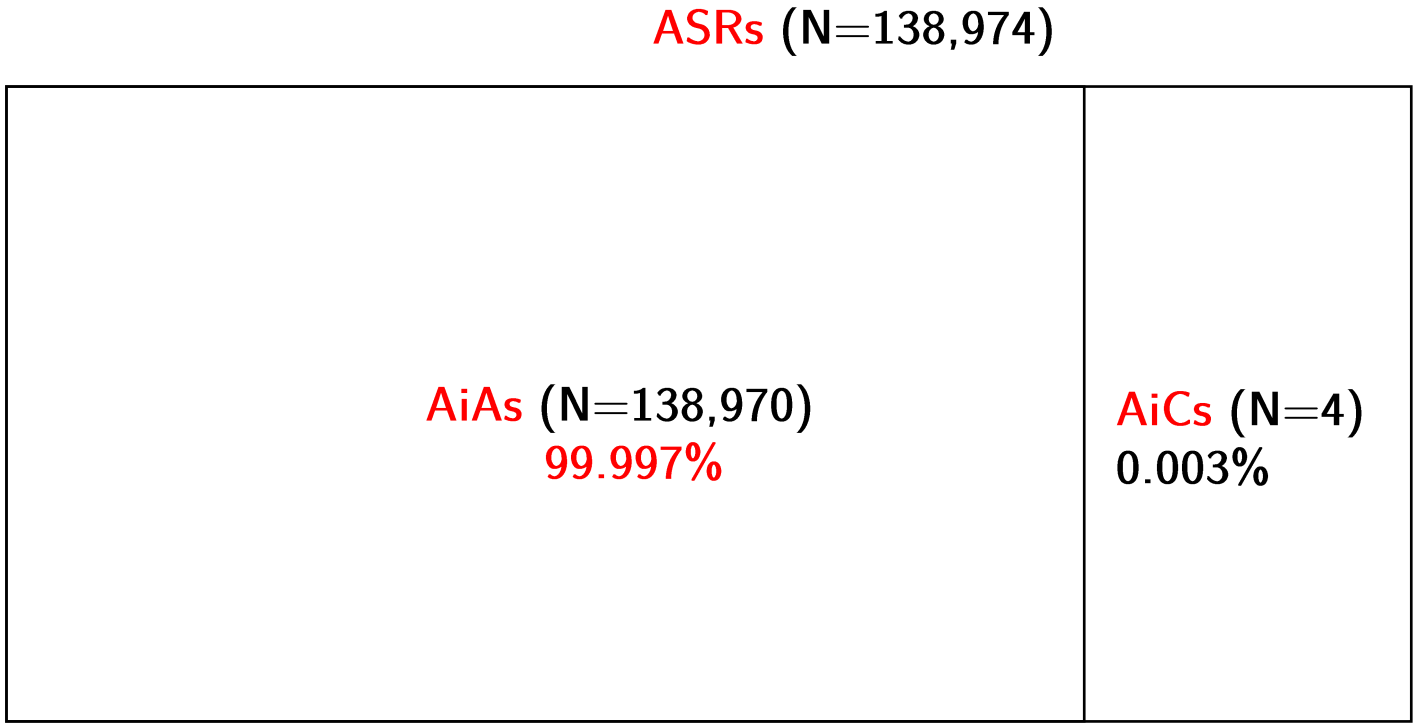
Big data analysis : descriptive study

- **ASRs** can be divided into two subgroups, 'Always-the-same-rating reviewers in All categories' **(AiAs)** and 'Always-the-same-rating reviewers in a category' **(AiCs)**.
- AiAs give the same star rating for all reviewed products in all categories,
- AiCs give the same star-rating for all reviewed products within one category.
- AiAs are therefore a subset of AiCs.

→ There are **138,974 unique ASRs** in all reviews (**138,974 AiAs** + **only 4 AiCs**).

Big data analysis : descriptive study

- There are 138,974 unique ASRs in all reviews and 138,970 of ASRs are AiAs and only four of ASRs are AiCs.

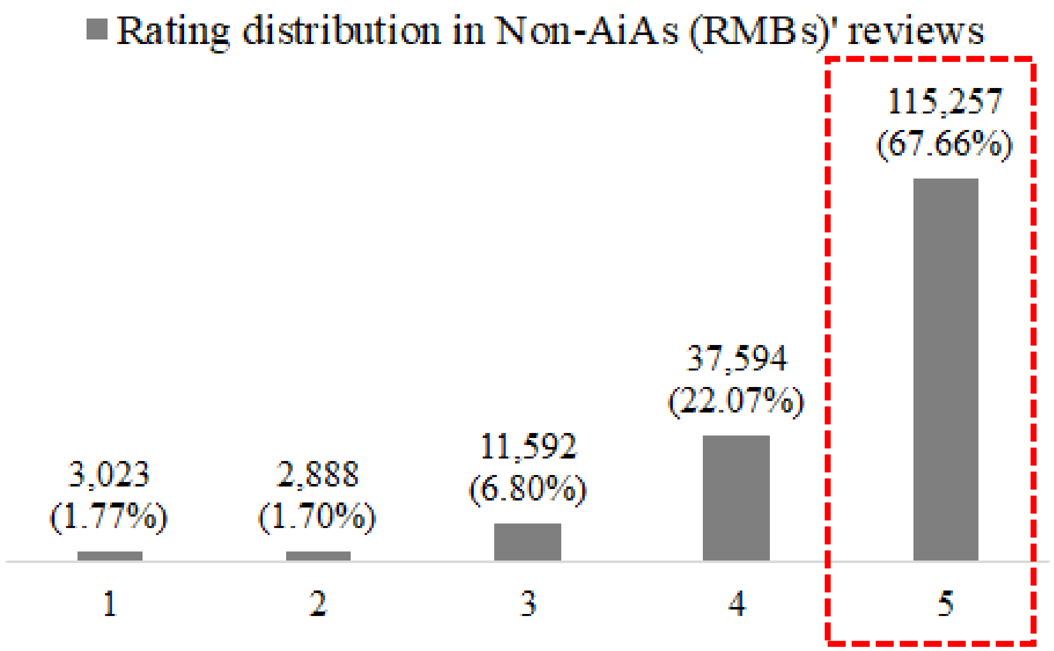


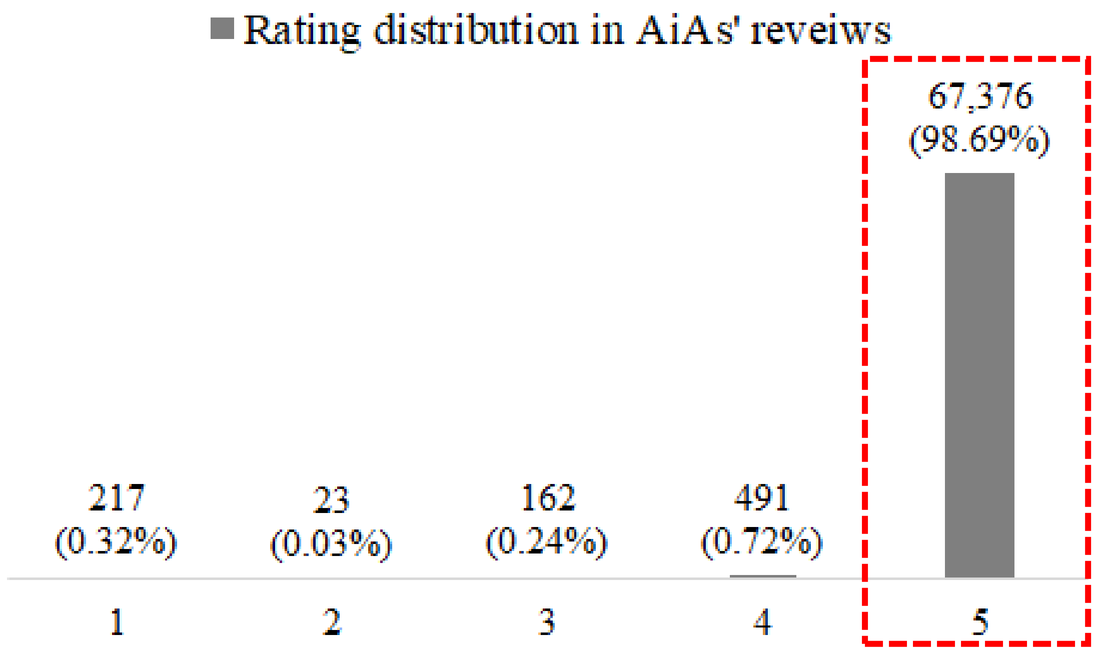
Big data analysis : descriptive study

Categories	Review	Reviewers	RMB (reviewers)	Alwaysers (reviewers)	ASR (reviewers)	ASR share in reviewers	ASR share in RMBs
Books	51,311,621	15,362,619	572,936	11,739,744	56,162	0.37%	9.80%
Clothing Shoes and Jewelry	32,292,099	12,483,678	281,349	9,150,747	23,131	0.19%	8.22%
Electronics	20,994,353	9,838,676	135,325	7,693,916	9,630	0.10%	7.12%
Home and Kitchen	21,928,568	9,767,606	143,232	7,560,187	11,650	0.12%	8.13%
Sports and Outdoors	12,980,837	6,703,391	62,188	5,491,623	5,636	0.08%	9.06%
Movies and TV	8,765,568	3,826,085	62,854	3,121,125	7,458	0.19%	11.87%
Cell Phones and Accessories	10,063,255	6,211,701	17,547	5,113,300	1,871	0.03%	10.66%
CDs and Vinyl	4,543,369	1,944,316	34,825	1,668,930	4,805	0.25%	13.80%
Kindle Store	5,722,988	2,409,262	48,939	2,012,255	5,409	0.22%	11.05%
Tools & Home Improvement	9,015,203	4,704,014	43,285	3,894,183	4,239	0.09%	9.79%
Toys and Games	8,201,231	4,204,994	42,058	3,498,007	7,029	0.17%	16.71%
Automotive	7,990,166	3,873,247	51,411	3,175,498	5,630	0.15%	10.95%
Pet Supplies	6,542,483	3,085,591	39,183	2,461,183	2,614	0.08%	6.67%
Office Products	5,581,313	3,404,914	14,028	2,968,653	2,182	0.06%	15.55%
Patio Lawn and Garden	5,236,058	3,097,405	14,236	2,640,191	1,479	0.05%	10.39%
Grocery and Gourmet Food	5,074,160	2,695,974	24,074	2,317,216	2,784	0.10%	11.56%
Video Games	2,565,349	1,540,618	8,997	1,325,081	1,129	0.07%	12.55%
Arts Crafts and Sewing	2,875,917	1,579,230	13,835	1,383,406	2,926	0.19%	21.15%
Musical Instruments	1,512,530	903,330	5,699	789,735	550	0.06%	9.65%
Digital Music	1,584,082	840,372	9,296	769,292	3,070	0.37%	33.02%
Industrial and Scientific	1,758,333	1,246,131	2,321	1,142,613	403	0.03%	17.36%
Software	459,436	375,147	262	351,048	7	0.00%	2.67%
AMAZON FASHION	883,636	749,233	89	704,353	7	0.00%	7.87%
Luxury Beauty	574,628	416,174	611	390,277	61	0.01%	9.98%
Appliances	602,777	515,650	71	496,663	38	0.01%	53.52%
All Beauty	371,345	324,038	11	315,174	2	0.00%	18.18%
Prime Pantry	471,614	247,659	2,787	217,091	595	0.24%	21.35%
Magazine Subscriptions	89,689	72,098	45	68,495	11	0.02%	24.44%
Gift Cards	147,194	128,877	26	127,620	18	0.01%	69.23%
Sum	230,139,802	102,552,030	1,631,520	82,587,606	160,526		

Big data analysis : descriptive study

- The **'digital music' category** is selected as a **target category** because it has a high share of ASRs among RMBs and the number of ASRs in the 'digital music' category is 3,070
- Rating distribution in RMBs' reviews in the digital music category

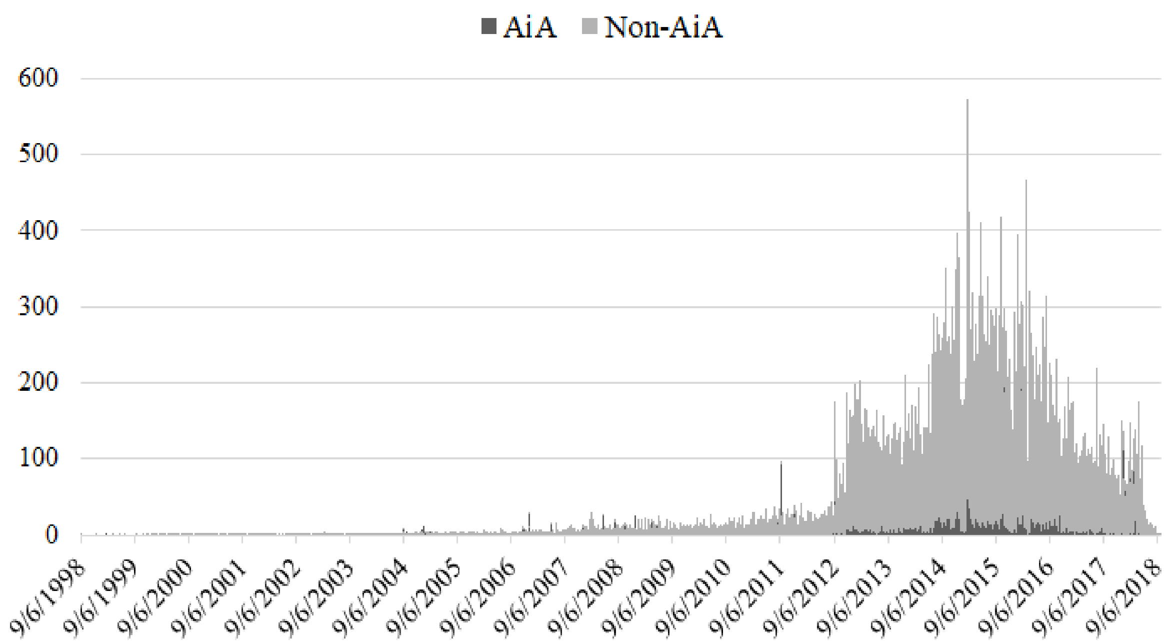




Big data analysis : descriptive study

- The number of RMBs' reviews in the digital music category over time

- The share of AiAs' reviews is the highest in 2015 as of 32.1%



	2014	2015	2016	2017
Non AiAs' reviews (A)	34,290	38,032	31,716	16,204
	70.5%	67.9%	68.9%	71.4%
AiAs' reviews (B)	14,381	**17,947**	14,311	6,499
	29.5%	**32.1%**	31.1%	28.6%
Total reviews (A+B)	48,671	55,979	46,027	22,703
Non-AiA reviewers (C)	3894	4089	3574	2375
	71.5%	70.8%	70.5%	71.9%
AiA reviewers (D)	1,549	1,689	1,495	929
	28.5%	29.2%	29.5%	28.1%
Unique reviewers (C+D)	5,443	5,778	5,069	3,304

3. Discrete choice analysis : descriptive study

Discrete choice analysis

- binary logistic models are applied to evaluate the determinant **AiAs' reviews** compared to **non-AiAs' reviews** between 2014 and 2017.
- purchase-verified AiAs and non-verified AiAs may differ in their tendencies to write reviews (Kim, Maslowska, and Malthouse 2018; Anderson and Simester 2014)
- Two questions are now explored.

→ What are the determinants of verified AiAs' reviews compared to verified non-AiAs' reviews?

→ What are the determinants of non-verified AiAs' reviews compared to non-verified non-AiAs' reviews?

Discrete choice analysis

- Distribution of Labels in Reviews

	AiAs' review dummy		
Verified dummy	0	1	Total
0 ('nvaia' models)	17,533 (model 1, y=0)	4,293 (model 1, y=1)	21,826
1 ('vaia' models)	102,709 (model 2, y=0)	48,845 (model 2, y=1)	151,554
Total	120,242	53,138	173,380

Discrete choice analysis

- Variables Description

Variable	Description
aia	AiAs' review dummy and base is non-AiAs' review (0)
aia_v	Verifiend AiAs' review dummy and base is verified non-AiAs' review (0)
aia_nv	Non-verifiend AiAs' review dummy and base is non-verified non-AiAs' review (0)
overall	i's star rating for reviewed digital music p at t_i
Vote	The number of helpfulness at t_i
summary_len	length of review summary (headline) at t_i
reviewtext_len	length of review body at t_i
u_n_review	i's number of reviews in digital category by t_i
u_n_rev_asin	i's number of repeated reviews for the digital music p by t_i
asin_n_rev	p's number of reviews by t_i
asin_avg_rating	p's average rating of reviews by t_i
asin_n_reviewers	p's number of reviewers posted reviews for p by t_i
asin_n_1_rating_share	p's share of 1-rating by t_i
asin_n_2_rating_share	p's share of 2-rating by t_i
asin_n_3_rating_share	p's share of 3-rating by t_i
asin_n_4_rating_share	p's share of 4-rating by t_i
asin_n_5_rating_share	p's share of 5-rating by t_i
asin_n_12_rating_share	p's share of 4-and 5-ratings by t_i (Consider correlation between 1 and 2 ratings)
asin_n_45_rating_share	p's share of 4-and 5-ratings by t_i (Consider correlation between 4 and 5 ratings)

Discrete choice analysis

- Empirical Results from Binary Logit Models

Variable	vaia_1	nvaia_1	vaia_2	nvaia_2	vaia_3	nvaia_3	vaia_4	nvaia_4
overall	2.871*** (0.055)	2.769*** (0.216)	2.812*** (0.062)	2.255*** (0.256)	2.847*** (0.052)	2.894*** (0.222)	2.846*** (0.053)	2.883*** (0.227)
vote	-0.006 (0.008)	-0.052*** (0.017)	-0.006 (0.008)	-0.054*** (0.018)	-0.007 (0.008)	-0.053*** (0.018)	-0.007 (0.008)	-0.055*** (0.018)
user_summary_len	-0.010*** (0.001)	0.001 (0.001)	-0.010*** (0.001)	0.001 (0.001)	-0.010*** (0.001)	0.001 (0.001)	-0.010*** (0.001)	0.001 (0.001)
user_reviewtext_len	-0.000*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)
u_n_review	-0.002*** (0.000)	-0.004*** (0.000)	-0.002*** (0.000)	-0.004*** (0.000)	-0.002*** (0.000)	-0.004*** (0.000)	-0.002*** (0.000)	-0.004*** (0.000)
u_n_rev_asin	0.193*** (0.033)	0.130** (0.060)	0.191*** (0.031)	0.147** (0.060)	0.193*** (0.032)	0.124** (0.060)	0.193*** (0.032)	0.130** (0.060)
asin_n_rev	0.001* (0.000)	0.004** (0.002)	0.001* (0.000)	0.004** (0.002)	0.001* (0.000)	0.003** (0.001)	0.001* (0.000)	0.004** (0.001)
asin_avg_rating	-0.088*** (0.021)	0.219*** (0.076)						
asin_n_reviewers	-0.001 (0.000)	-0.004** (0.002)	-0.001 (0.000)	-0.004*** (0.002)	-0.001 (0.000)	-0.004** (0.002)	-0.001 (0.000)	-0.004** (0.002)
holiday	-0.145*** (0.040)	0.041 (0.129)	-0.145*** (0.040)	0.041 (0.130)	-0.145*** (0.040)	0.047 (0.129)	-0.145*** (0.040)	0.054 (0.129)
asin_n_1_rating_share			0.846*** (0.148)	3.530*** (0.806)				
asin_n_5_rating_share			0.071 (0.046)	1.446*** (0.200)				
asin_n_4_5_rating_share					-0.457*** (0.069)	0.204 (0.277)	-0.422*** (0.103)	1.694*** (0.433)
asin_n_1_2_rating_share							0.070 (0.155)	2.609*** (0.707)
N	151,554	21,826	151,554	21,826	151,554	21,826	151,554	21,826
Log Likelihood	-84,050.204	-8,526.120	-84,041.246	-8,471.087	-84,037.652	-8,530.611	-84,037.55	-8,519.001
AIC	168,162.41	17,114.241	168,146.49	17,006.173	168,137.3	17,123.223	168,139.1	17,102.001
BIC	168,470.2	17,361.957	168,464.21	17,261.881	168,445.09	17,370.939	168,456.82	17,357.709

Discrete choice analysis

- The empirical findings of the binary logit model suggest that:

→ the reviews that contain a higher average rating are less likely to be verified AiAs' reviews,

but are more likely to be non-verified AiAs' reviews.

→ Increasing the share of positive ratings (4- and 5- star ratings) in reviews of the given digital music

decreases the probability that the review has been written by a verified AiA

increase the probability that the review has been written by a non-verified AiA.

→ Increasing the share of negative ratings (1- and 2-star ratings) in reviews of the given digital music also

increases the probability that the review has been written by a non-verified AiA instead of a non-verified non-AiA.

Discrete choice analysis

- the probability that the review has been written by a AiA regardless of purchase verification increases with

(1) higher ratings from a reviewer of digital music,

(2) shorter review texts,

(3) a smaller number of reviews from the reviewer,

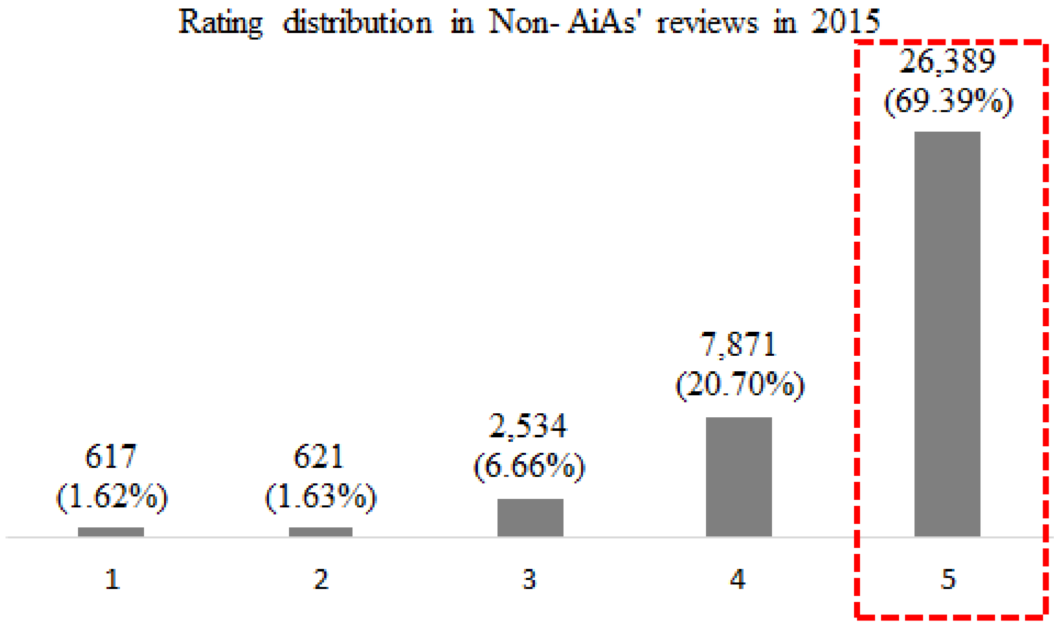
(4) a higher number of repeated reviews from the reviewer of the given digital music, and

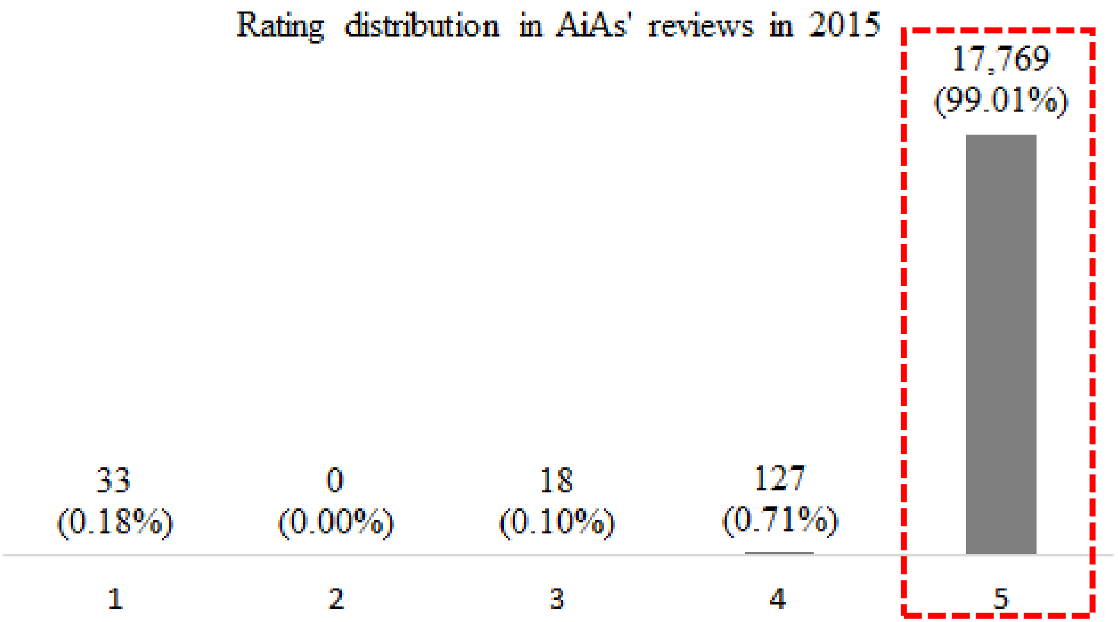
(5) a larger number of reviews for digital music.

4. Review classification using AI : digital experiment

Review classification using AI

- Rating distribution in Non-AiAs and AiAs' reviews during 2015





Review classification using AI

- Distribution of Non-AiAs and AiAs and their reviews in each dataset

	Total Set		Total Training Set		Training Set		Valid Set		Test Set	
	Count	Shares	Count	Shares	Count	Share	Count	Share	Count	Share
Non AiAs' reviews	38,032	67.94%	30,392	67.87%	22,657	67.46%	7,735	69.09%	7,640	68.24%
AIAs' reviews	17,947	32.06%	14,391	32.13%	10,930	32.54%	3,461	30.91%	3,556	31.76%
Total	55,979	100.00%	44,783	100.00%	33,587	100.00%	11,196	100.00%	11,196	100.00%
Non AiAs	4,089	70.77%	3,706	71.11%	1,286	28.50%	608	25.87%	617	27.13%
AIAs	1,689	29.23%	1,506	28.89%	3,227	71.50%	1,742	74.13%	1,657	72.87%
Total reviewers	5,778	100.00%	5,212	100.00%	4,513	100.00%	2,350	100.00%	2,274	100.00%
Period	1/1/2015 – 2015-12-31		1/1/2015 – 2015-10-07		1/1/2015 -2015-07-23		7/23/2015 -10/7/2015		10/7/2015 -12/31/2015	

Review classification using AI

- Research design

Models	Classifier	Feature sets	Word embedding
Base model	Logistics regression	observational variables only	N/A
Parital deep learning	CNN*	Text only	pre-trained BERT
	Weighted* CNN	Text only	pre-trained BERT
Full depp learning	CNN	observational data + Text	pre-trained BERT
	Weighted CNN	observational data + Text	pre-trained BERT

Review classification using AI

- Positive weight is applied into binary cross-entropy loss function to mitigate the imbalanced problem in this dataset as follow:

$$\text{Loss}\,(x_i, \mathrm{y}_i) = -\left[\text{positive} - \text{weight} \times \mathrm{y}_i \log\left(\frac{1}{1+\mathrm{e}^{-x_i}}\right) + (1-\mathrm{y}_i)\log\left(\frac{1}{1+\mathrm{e}^{-x_i}}\right)\right]$$

- Where positive weight is $\frac{\mathrm{N}}{\mathrm{K} \times \mathrm{N_p}}$, and N is the number of the sample; K is the number of classes; and, $\mathrm{N_p}$ is the number of sample belong to positive class.

Review classification using AI

- Base model (only non-textual variables)

Models	Word Embedding	Hyperparameter	Accuracy	Precision	Recall	F1-score
Logitic regression	N/A	N/A	.675	1: .69 2: .45 WA: .61	1: .95 2: .10 WA: 0.68	1: .80 2: .16 WA: **0.60**

Review classification using AI

- The Prediction results of the partial models (text-only) for AiAs' reviews classification

Models	Word Embedding	Hyperparameter	Accuracy	Precision	Recall	F1-score
CNN	BERT	Max length: 512 Epoch: 1 Number of filters: 200 Filter sizes: (3,4,5) Dropout: 0.7 Learning rate: 0.00001	.682	1: .68 2: .00 WA: .47	1: 1.00 2: .00 WA: .68	1: .81 2: .00 **WA: .55**
Weighted CNN	BERT	Max length: 512 Epoch: 3 Number of filters: 200 Filter sizes: (3,4,5) Dropout: 0.7 Learning rate: 0.0001 Positive weighted: 1.536	.659	1: .69 2: .40 WA: .60	1: .90 2: .15 WA: .66	1: .78 2: .22 **WA: .60**

Review classification using AI

- Prediction results of the full models for AiAs' reviews classification

Models	Word Embedding	Hyperparameter	Accuracy	Precision	Recall	F1-score
CNN	BERT	Max length: 512 Epoch: 7 Number of filters: 300 Filter sizes: (2,3,4) Dropout: 0.6 Learning rate: 0.0001	.651	1: .70 2: .41 WA: .61	1: .84 2: .24 WA: .65	1: .77 2: .30 WA: **0.62**
Weighted CNN	BERT	Max length: 512 Epoch: 5 Number of filters: 200 Filter sizes: (2,3,4) Dropout: 0.6 Learning rate: 0.0001 Positive weighted: 1.536	.640	1: .72 2: .42 WA: .62	1: .77 2: .35 WA: .64	1: .75 2: .38 WA: **.63**

5. Conclusion & Contribution

Conclusion : big data analysis

- Surprisingly, some reviewers always write the same star rating for all reviewed products in a category or all the categories.
- These always-the-same rating reviewers (ASRs) are Always the same rating reviewers in all categories (AiAs) in 99.99% excluding only 4 reviewers who always the same rating reviewers in a category (AiCs).
- In addition, **most AiAs are always-happy reviewers (AHRs) who always give five-star ratings for reviewed products**.
- These points indicate that the reviews written by **ASRs might cause an upward bias for product quality estimation**.

Conclusion : discrete choice analysis

- This study empirically demonstrates that **star rating, the usefulness of reviews, and length of the headline and review** are potential indicators of reviews written by ASRs.
- In particular, the main difference between **verified** and **non-verified AiAs reviews** are **the effect of average star ratings and extreme star ratings**.

Conclusion : Classification

- **The positive weighted CNN on top of BERT** embedding shows higher predictive performance in the F1 score than the unweighted CNN on top of BERT embedding.

- Further, **combining text and non-text data shows a higher performance than using only text data**. This point shows the potential for deep learning to detect biased reviews by using text and non-textual variables.

Contribution

- Some studies have applied causal inference methods, such as regression discontinuity design (RDD) and difference-in-difference (DiD) to examine the effects of online product reviews on sales.
- The approaches in this study might be useful for mitigating **the effects of potential self-selection bias** in online reviews before the application of causal inference methods.

Q&A

jikhan.jeong@wsu.edu