Automatically Discovering Unknown Product Attributes Impacting Consumer Preferences

Ankit Sisodia, Alex Burnap, Vineet Kumar
Yale School of Management, ankit.sisodia@yale.edu, alex.burnap@yale.edu, vineet.kumar@yale.edu

February 2021
Latest version at:
http://vineetkumars.github.io/Papers/AutomaticAttributeDiscovery.pdf

Marketers use quantitative models of consumer preferences over products to understand and predict market demand. Conventionally, this requires the marketer to *manually* define which product attributes are relevant to model for the consumer. However, it is not always easy to understand from observational data which product attributes actually drive consumer demand. In this work, we aim to *automatically* discover attributes using a combination of machine learning and econometrics. We build on recent advances in disentangled representations to learn independent factors of variation–discovered attributes-in large-scale unstructured data (e.g., images). These discovered attributes are jointly inferred within an econometric model to simultaneously quantify their impact on an economic primitive (e.g., willingness to pay (WTP). Our approach does not rely on marketer judgment to discover attributes, yet outputs attributes that are semantically interpretable and statistically independent. This ensures discovered attributes are both economically important and interpretable to managers. We apply this approach to automatically discover visual product attributes of high-end watches auctioned at Christie's, in which the auction format ensures truthful estimates of WTP to permit causal identification of discovered attributes' economic impact. We evaluate discovered attributes along their semantic interpretability, statistical independence, and ability to explain variation within the econometric model. Our preliminary results for all three evaluation criteria indicate the proposed method has promise for automatic discovery of attributes that impact consumer preferences.

Key words: deep learning, disentanglement, visual analytics

1. Introduction

Marketing uses quantitative models to understand and predict consumer purchase behavior based on consumer characteristics and product attributes. In these models, defining the relevant set of attributes that matter to the consumer is tasked to the marketer, and is often driven by availability of structured data. For many products, the set of attributes may be obvious or well known to managers, e.g in consumer packaged goods. However, for other categories, e.g. cars or watches, there may be several nuanced attributes that are not *a priori* obvious to the modeler, but significantly impact consumer preferences. Product design can impact much more than just aesthetics, including the functionality and usability of a product. To quote Steve Jobs: "Design is not just what it looks like and feels like. Design is how it works."

In this paper, we propose a method to automatically discover unknown product attributes from unstructured data (e.g., images) that can determine an economic primitive of interest (e.g., willingness to pay or demand). Our proposed method can help both researchers and managers alike discover product attributes that are hard to define or even unknown. This helps avoid modeling errors that can lead to biased or even contradictory results. For managers, including these now discovered attributes within models can, for example, improve managerial understanding of how their product compares with competing products along attributes that impact consumer preferences. For researchers, since our method both discovers unknown attributes and their values, our method can be used as an input to other models that investigate substantive marketing questions. We show an example of the latter with an application in the visual domain, in which our method is used to discover visual product attributes of luxury watches that drive consumers' willingness to pay.

Our proposed approach lies at the intersection of deep learning and econometrics. This gives our method a combination of several advantages when compared with existing marketing and econometric methods. First, it discovers attributes automatically from unstructured data without human specification of how many attributes there are, or what those attributes might be. In contrast, conventional methods require researcher judgment to define product attributes and then human labeling to give values to those attributes. Second, it aims to discover attributes that are statistically independent of each other, yet semantically interpretable by humans. This gives modeling flexibility as well as managerial practicality. Third, our method jointly infers both the discovered attributes as well as the dependence of economic primitives on these discovered attributes.

Existing marketing methods decouple the process of discovering attributes and their dependence on economic primitives as separate sequential processes. Our method has several advantages over the current state-of-the-art. First, this decoupled process can lead to missing out many relevant product attributes, whereas our method is able to obtain multiple important attributes. Second, in traditional models, the researcher's selection of which variables to include might bias the inference of the economic primitive. Third, there might be undiscovered variables correlated with the researcher's selected variable that have a greater true impact on the economic primitive. Collectively, these might lead to misleading conclusions about the importance of selected variables. In contrast, our method does not rely on researcher selection of variables from unstructured data.

We build on recent advances in *disentanglement*, an area of deep learning. In this literature, deep neural networks are used to learn representations of unstructured product data (e.g., images, text, or videos) as substantially lower dimensional but semantically meaningful latent attributes. We aim to use *disentanglement* to separate and obtain distinct yet semantically meaningful product attributes while simultaneously learning their connection to an economic primitive of interest. Our implementation is based on variational autoencoders (VAEs), such that our method includes both an "encoder" and "decoder." The encoder takes unstructured data as an input and outputs latent low-dimension attributes, whereas the decoder attempts to reconstruct the original data from the low-dimensional attributes. Our "encoder" is additionally connected to an econometric model tying the discovered attributes to an economic primitive of interest (e.g., WTP).

We show a preliminary application in the visual domain. Specifically, our aim is to discover visual attributes of luxury watches and understand their impact on consumer's willingness to pay. Our data is comprised of several brands of luxury watches auctioned at Christie's from 2001 to 2020. This application is chosen for two reasons: (1) Visual product attributes are important drivers of consumer demand and thus of considerable interest to marketers. These attributes can be a source of horizontal differentiation in product categories with functional similarity between products. In other categories, they can be a source of vertical differentiation, and allow firms to charge a premium for products with preferred design or visual attributes. (2) The auction mechanism for this data results in Hammer prices, a true estimate of consumers' willingness to pay, and thereby an econometric model that requires minimal assumptions.

Our method automatically discovers visual attributes that can be semantically interpreted as dial color, strap color, dial shape, and the ratio of dial size to strap length. We also infer values of these discovered attributes for all watches in our dataset. Our method did not require any human intervention in this task. We demonstrate that the discovered attributes are independent and semantically interpretable. We recover the distribution of all of the discovered attributes and find 6 attributes to be semantically meaningful. We then measure the impact of the recovered visual attributes on consumer's willingness to pay. We find that model fit and predictive ability on out of sample data is greatest when the discovered visual attributes are combined with structured attributes. After controlling for the watch brand and auction attributes (data and location of the auction), we find that overall size of the watch followed by the ratio of dial size to strap length are the two most important drivers of the hammer price.

There are a few limitations of our approach that are worth mentioning. First, it requires an economically meaningful variable like demand or willingness to pay, and requires that it be matched with the unstructured data so they can be connected to assess the dependence. Another limitation of our method is that it requires data to be pre-processed. The watches in our dataset were all vertically aligned in orientation, were centered and had white background color. This preprocessing allows our method to only recover economically meaningful visual attributes, and is a one-time task before the algorithm can be run on the data.

2. Literature Review

Our work is related to three broad streams of literature. First, it is related to marketing methods for attribute discovery. Second, it is related to marketing and economics methods to investigate the effect of product attributes on economic primitives of interest such as consumer demand and consumer willingness to pay. Third, our work relates to a stream of literature in machine learning known as representation learning and more specifically to disentangled representations.

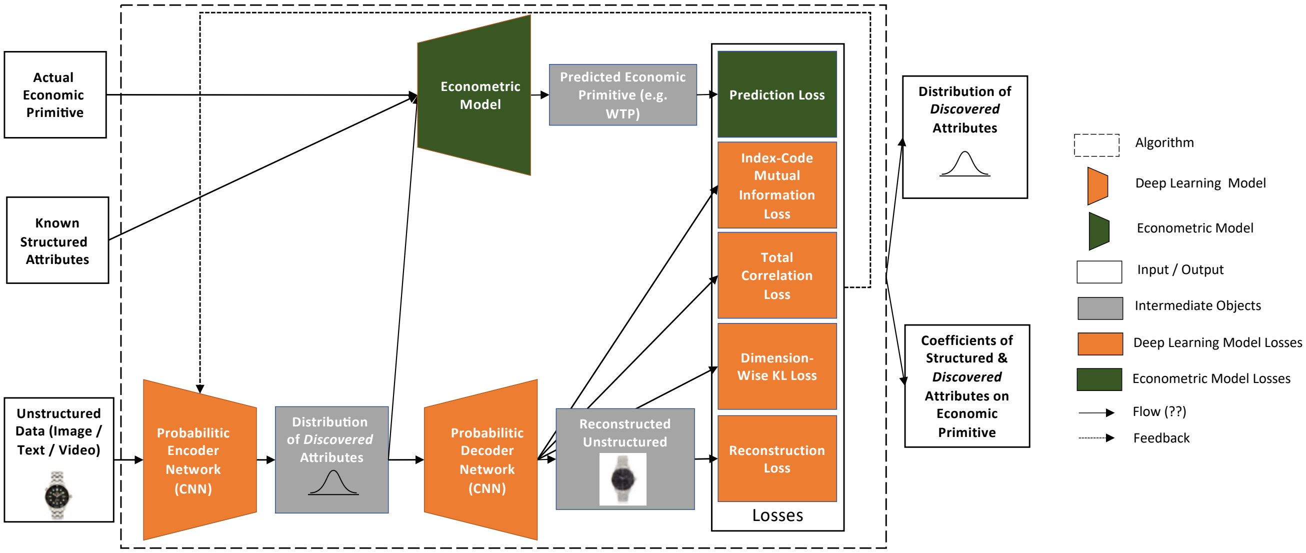
Conventional methods in marketing science require a defined list of product attributes over which consumers form preferences as inputs. Examples of methods that need a set of attributes as inputs range from conjoint analysis and factor analysis, to reduced-form regression models and structural models. Along with the product attributes, these methods also need attribute levels. Researchers have relied on focus groups, in-depth consumer interviews and internal corporate expertise to structure the set of attributes and levels for use in these methods. (Green et al. 2001). This approach suffers from two drawbacks. First, it requires extensive human intervention in both defining the set of attributes as well as their levels. Second, it provides no guarantee to select attributes that could be correlated with an economic primitive of interest as they might be non-obvious ex-ante and thus, hard to recover from market research. Our approach can complement existing market research methods to discover additional independent product attributes automatically from unstructured data that can also be semantically interpreted.

Our work also relates to models in marketing and economics in which an economic primitive is assumed to be dependent on consumer's preferences over a set of attributes as well as price. Examples of such models include those derived from the traditions of scanner panel data literature in marketing (Guadagni and Little 1983). The class of demand models inspired by Berry 1994, Berry et al. 1995 (BLP) were among the first to include an unobservable quality (or "structural error") that impacted demand for a product. Our approach has a number of differences. First, while the BLP class of models allow for estimation of unobserved quality, there is only one dimension to capture both visual and any other (e.g. experience) set of attributes. The implication is that we cannot attach any product-level interpretability to this factor. Second, our approach recovers the distribution of multiple attributes along with their attribute levels from unstructured data. This attribute set can be fed into models of individual or aggregate demand to answer more substantive counterfactual questions such as the effect on market outcomes obtained by changing a particular attribute while holding others invariant.

Finally, our work relates to a stream of literature in machine learning known as representation learning. The goal of representation learning is to learn efficient coding of data for knowledge transfer and generalization outside the original representation of the data. Disentangled representations in particular aim to additionally separate the distinct informative factors of variation in the data (Bengio et al. 2013). Several recent approaches for learning disentangled representations are based on Variational Autoencoders (VAEs) (Kingma and Welling 2014), including β-VAE (Higgins et al. 2017), AnnealedVAE (Burgess et al. 2018), FactorVAE (Kim and Mnih 2018), and β-TCVAE (Chen et al. 2018).

Our work puts an additional constraint on these recent deep learning methods in that the discovered factors of variation (the "discovered" product attributes in this paper) must also correlate with economic primitives of interest within an econometric model. This helps ensure discovered product attributes are not only independent and semantically interpretable, but that they are also economically meaningful. From this viewpoint, while these recent deep learning methods are generally aimed at learning disentangled representations from high-dimensional unstructured data (e.g., images, text, video), they may also be viewed as high-dimensional and nonlinear extensions of classic marketing methods such as (supervised) factor analysis and independent component analysis, in which the learned representations are statistically independent; albeit much lower-dimensional and obtained using linear projections (Roweis and Ghahramani 1999).

Figure 1 Schematic Illustration of Proposed Approach


Actual Economic Primitive
Known Structured Attributes
Unstructured Data (Image / Text / Video)
Econometric Model
Predicted Economic Primitive (e.g. WTP)
Probabilitic Encoder Network (CNN)
Distribution of *Discovered* Attributes
Probabilitic Decoder Network (CNN)
Reconstructed Unstructured
Prediction Loss
Index-Code Mutual Information Loss
Total Correlation Loss
Dimension-Wise KL Loss
Reconstruction Loss
Losses
Distribution of *Discovered* Attributes
Coefficients of Structured & *Discovered* Attributes on Economic Primitive
Algorithm
Deep Learning Model
Econometric Model
Input / Output
Intermediate Objects
Deep Learning Model Losses
Econometric Model Losses
Flow (??)
Feedback

3. Methodology

Our proposed approach integrates machine learning and econometrics methods. We build on recent methods in disentangled representation learning using Variational Autoencoders (VAEs) (Kingma and Welling 2014), including β-VAE (Higgins et al. 2017), Annealed-VAE (Burgess et al. 2018), FactorVAE (Kim and Mnih 2018) and β-TCVAE (Chen et al. 2018). By integrating disentangled representation learning with econometric methods, we aim to discover hitherto unknown attributes that are independent, can be semantically interpreted, and are economically relevant. Moreover, we aim to discover these attributes automatically, in a manner that does not require researcher judgment as is commonly required with conventional marketing methods.

Figure 1 represents an an illustration of our approach. The model receives unstructured data (e.g., image), its associated known structured attributes (e.g., brand), and its associated economic primitive (e.g., willingness to pay) as input. Then, our model *discovers* the unknown attributes along with their distribution parameters.

The model is trained to minimize the weighted sum of the following five losses below.

1. *Reconstruction Loss* represents the difference between the reconstructed unstructured data (text or image) and the original unstructured data. It ensures that the learned low dimensional attributes possess the necessary information to be able to reconstruct the product image.
2. *Prediction Loss* represents the difference between the actual and the predicted willingness to pay is as low as possible. It ensures that the discovered attributes are economically relevant as they impact the economic primitive of interest.
3. *Index-Code Mutual Information Loss* represents the mutual information shared between the discovered attributes and the original data. It ensures the model is able to discard nuisance attributes.
4. *Total Correlation Loss* represents the correlation between discovered attributes, as characterized by the distance between the joint distribution of these attributes compared to their marginal distributions taken separately.
5. *Dimension-Wise KL Loss* ensures the distance between the distribution of the discovered attributes is close to the specified prior distribution.

3.1. Two-Step Generative Process for Unstructured Data and Economic Primitive

We next describe the method in detail. Our description focuses on the case where the unstructured data are visual images, discovered attributes are independent low-dimensional factors characterizing those images, and the economic primitive of interest is consumer willingness to pay (WTP). Let us consider the dataset $\mathbf{X}, \mathbf{Y}, \mathbf{V}, \mathbf{A} = \{(\mathbf{x}_1, y_1, \mathbf{v}_1, \mathbf{a}_1), ..., (\mathbf{x}_N, y_N, \mathbf{v}_N, \mathbf{a}_N)\}$ where the i-th observation is a product image $\mathbf{x}_i \in \mathbb{R}^D$, its corresponding willingness to pay $y_i \in \mathbb{R}$, its corresponding structured attributes $\mathbf{v}_i \in \mathbb{R}^K$ and auction attributes $\mathbf{a}_i \in \mathbb{R}^L$.

Our model assumes a two-step data generating process. The first step samples the (unknown and discoverable) visual attributes and known structured attributes, while the second step uses these attributes to generate a product image $\mathbf{x}_i$ and its willingness to pay y_i. Formally, the first step samples the unknown visual attributes denoted by $\mathbf{z}_i \in \mathbb{R}^J$ is sampled from a centered isotropic multivariate Gaussian $p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$ so that it is completely factorised. Note that we need to tune the dimensions of the latent visual attributes J in the model, as the choice J of the latent space of unknown visual attributes involves an important tradeoff. With a small J, the model might combine multiple visual attributes into one which results in entanglement; however, when J is large, the model discovers redundant or irrelevant attributes or it might even break up a true attribute across multiple dimensions. In the second step, the reconstructed product image $\mathbf{x}_i$ is generated from the conditional distribution $p_\theta(\mathbf{x}|\mathbf{z})$ as parametrized using a multivariate Gaussian. Likewise, the consumer's willingness to pay y is generated from the conditional distribution $p_w(y|\mathbf{z},\mathbf{v},\mathbf{a})$. We refer to $p_\theta(\mathbf{x}|\mathbf{z})$ as a probabilistic decoder and the $p_{\mathbf{w}}(y|\mathbf{z},\mathbf{v},\mathbf{a})$ as the econometric model.

As standard in variational Bayesian inference (Blei et al. 2017), the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$ is intractable. We adopt the common VAE assumption that the true posterior can instead be approximated using a variational family of Gaussians with diagonal covariance $\log q_\phi(\mathbf{z}|\mathbf{x}) = \log \mathcal{N}(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma}^2\mathbf{I})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the mean and the s.d. of the approximate posterior. Note that $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the parameters of each observation i's distribution of discovered visual attributes. The econometric model can use the information from the entire distribution of the *discovered* attributes in addition to the usual structured attributes. This allows far more modeling flexibility as compared to typical point estimates. We refer to $\log q_\phi(\mathbf{z}|\mathbf{x})$ as the probabilistic encoder whose parameters are jointly learned along with the parameters of the probabilistic decoder using a neural net.

We train the model by minimizing a variational bound to the negative log-likelihood, and use the backpropagation algorithm to find values of the parameters $(\theta,\phi,\mathbf{w})$ of the deep learning networks that minimize the bound. The parameter space or number of weights of the deep neural networks in our application are often in the range of hundreds of thousands depending on architectural decisions (our chosen architecture in Section 4 has 504,066 parameters). The variational bound combined with the econometric model results in a total loss function given in Equation (1).

$$
\begin{aligned}
\underbrace{L(\theta,\phi,\mathbf{w})}_{\text{Total Loss}} = {} & \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]}_{\text{Reconstruction Loss}} + \delta \underbrace{MSE(\hat{y}(\mathbf{z},\mathbf{v},\mathbf{a}),y)}_{\text{Prediction Loss}} \\
& + \underbrace{\alpha \quad I_q(\mathbf{z},\mathbf{x})}_{\text{Index-code Mutual Information Loss}} + \underbrace{\beta KL\left[q(\mathbf{z})||\prod_{j=1}^{J} q(z_j)\right]}_{\text{Total Correlation Loss}} + \underbrace{\gamma \sum_{j=1}^{J} KL\left[q(z_j)||p(z_j)\right]}_{\text{Dimension-Wise KL Divergence Loss}}
\end{aligned}
\tag{1}
$$

The loss minimization problem spans over the set of parameters $(\theta, \phi, \mathbf{w})$ where θ and ϕ are the parameter vectors of the deep neural network using unstructured data, and $\mathbf{w}$ is the parameter vector of the econometric model.

The terms in this equation are the reconstruction loss, prediction loss, index-code mutual information loss, total correlation loss, dimension-wise KL-divergence loss, respectively. The index-code mutual information loss, the total correlation loss, and the dimension-wise KL divergence loss act as regularization losses in the VAE literature (Kim and Mnih 2018, Chen et al. 2018). Penalizing the reconstruction loss encourages the reconstructed output to be as close as possible to the input data. $I_q(\mathbf{z}, \mathbf{x})$ is the mutual information between the visual attribute $\mathbf{z}$ and the product image $\mathbf{x}$. Penalizing this term reduces the amount of information about $\mathbf{x}$ stored in $\mathbf{z}$. A low α stores nuisance information in $\mathbf{z}$ while a high α leads to poor reconstruction. A negative α would encourage higher mutual information between the visual attribute $\mathbf{z}$ and the product image $\mathbf{x}$. The Total Correlation loss denoted as $KL\left[q(\mathbf{z})||\prod_{j=1}^{J} q(z_j)\right]$ represents a measure of dependence of multiple random variables in information theory. Penalizing the total correlation term forces the model to find statistically independent visual attributes. A high β results in a more disentangled representation but with poor reconstruction. $\sum_{j=1}^{J} KL\left[q(z_j)||p(z_j)\right]$ is referred to as dimension-wise KL divergence term. Penalizing this term pushes $q(z_j)$ to the prior $p(z_j)$ encouraging probabilistic structure imposed by the parametric assumptions of the prior (e.g., Gaussian). Penalizing the mean-squared error prediction loss $MSE(\hat{y}(\mathbf{z}, \mathbf{v}, \mathbf{a}), y)$, where $\hat{y}(\mathbf{z}, \mathbf{v}, \mathbf{a}) \sim p_{\mathbf{w}}(y|\mathbf{z}, \mathbf{v}, \mathbf{a})$ prioritizes the structured and discovered visual attributes $\mathbf{z}$ to obtain high accuracy in predicting y.

3.2. Economically-Meaningful Supervision for Disentangled Representations

Our primary methodological contribution is economically-meaningful supervision for otherwise unsupervised methods of disentangled representation learning. Consider in contrast an alternative approach in which an unsupervised disentanglement algorithm obtains discovered attributes, followed sequentially by their input into a classical econometric model (e.g. OLS regression) to determine an economic primitive (e.g. willingness to pay). This alternative approach clashes with impossibility theorems (see e.g., Locatello et al. 2019) that prove that one cannot ensure discovered visual attributes are disentangled with only an unsupervised algorithm.

It is helpful to detail the intuition here. If an unsupervised method learns a generative model so that there are perfectly disentangled visual attributes of images, then then there is an equivalent generative model where the visual attributes are completely entangled. From an econometric standpoint, there exist an infinite number of identical representations of the "discovered" visual attributes $\mathbf{z}$, all of which can generate the same data $\mathbf{x}$. Complete entanglement implies that a single change in a single discovered visual attribute results in changes to *all* the visual attributes. Now, since an unsupervised algorithm only has access to images, it won't be able to distinguish between the two equivalent generative models.

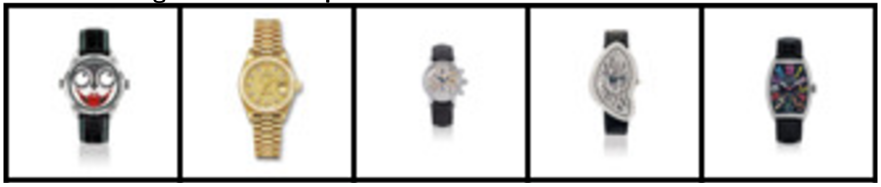
In our approach, we use an economic primitive associated with the unstructured data for supervision. We overcome the impossibility theorem results by imposing constraints that the discovered disentangled visual attributes impact the economic primitive of interest. By correlating the visual attributes with an economic primitive, we not only overcome the impossibility theorem results but also ensure that discovered distribution of the visual attributes are also economically meaningful. In sum, our approach *simultaneously* learns to discover unknown product attributes and their distribution, quantifies their levels for each product and recovers the impact of visual attributes on consumer willingness to pay, Moreover, the discovered attributes are independent of each other and can be semantically interpreted.

4. Application to Visual Domain: Watch Auctions

We provide an application of our proposed method in the visual domain. Examples of visual domains in marketing that impact consumer demand include a product's image, product's packaging and product's promotion materials. Understanding their impact on consumer demand is of considerable interest (Kang et al. 2019, Burnap et al. 2019). Existing methods treat either ignore visual attributes completely, or aggregate across all visual (and other) unobservable attributes to form an unobserved product characteristic (Cho et al. 2015).

Using our method, we discover independent visual attributes that can be semantically interpreted as well as their values. Then, we characterize the effect of each discovered visual attribute along other known structured attributes on consumer demand. While our method can be applied across multiple modes of data with unobserved attributes (e.g. text, audio, video), visual attributes are a natural, substantively important category that also involves technical challenges. On the one hand, it is hard to dispute the importance of product aesthetics and design on consumer demand (Dotson et al. 2019). At the same time, it is challenging to define visual attributes of a product category and to quantify their levels.

Figure 2 Example of Watches Auctioned at Christie's



4.1. Data and Empirical Setting

We apply our method to discover the unknown visual attributes of luxury watches, which are regarded as collectibles and sold through auctions. We choose this data for two main reasons. First, visual attributes of watches are important considerations for consumers. Second, the mechanism of auction leads to a truthful revelation of the buyer's willingness to pay (WTP) for the watches. We use data from watches auctioned at the Christie's auction house, spanning the years 2001 to 2020.

For each auctioned watch in the dataset, we have its image, auction attributes, structured product attributes, and the hammer price paid at the auction (i.e., the willingness to pay). Structured attributes are the brand of the watch, year of manufacture or *circa*, type of movement associated with the watch, dimensions of the watch and materials used in the watch. Auction attributes are year of the auction, location of the auction. Figure 2 shows a sample of watch images in our dataset. The hammer price corresponding to a consumer's willingness to pay (in $1000s) are in 2000 dollars, adjusted for inflation using the Consumer Price Index. Figure 5 shows the brand-wise distribution of the hammer price. The auction location was both online as well as offline. The offline locations are Amsterdam, Dubai, Hong Kong, London and New York. A total of 199 unique brands are present in the data. Audemar's Piguet, Cartier, Patel Philippe and Rolex are the four brands with the largest share of observations. Circa is coded as Pre-1950, 1950s, 1960s, 1970s, 1980s, 1990s, 2000s and 2010s. Movement of a watch is classified as either mechanical, automatic or quartz. Dimensions of the watch means the watch diameter in case of a circular dial or the length of the longest edge in case of a rectangular dial (in millimeters). We include dummy variables to denote whether a watch material includes ceramic, diamond, gold, platinum, steel or titanium. Table 1 provides summary statistics of the auctioned watches. Figure 3 and Figure 4 provide the distribution of the hammer price on structured and auction attributes.

Table 1 Summary Statistics of Auctioned Watches

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Auction Location (Amsterdam)	6,830	0.006	0.075	0	0	0	1
Auction Location (Dubai)	6,830	0.124	0.329	0	0	0	1
Auction Location (Hong Kong)	6,830	0.362	0.481	0	0	1	1
Auction Location (London)	6,830	0.024	0.153	0	0	0	1
Auction Location (New York)	6,830	0.204	0.403	0	0	0	1
Auction Location (Online)	6,830	0.281	0.450	0	0	1	1
Brand (Audemar's Piguet)	6,830	0.061	0.239	0	0	0	1
Brand (Cartier)	6,830	0.073	0.260	0	0	0	1
Brand (Patek Philippe)	6,830	0.187	0.390	0	0	0	1
Brand (Rolex)	6,830	0.180	0.384	0	0	0	1
Brand (Others)	6,830	0.498	0.500	0	0	1	1
Circa (Pre-1950)	6,830	0.047	0.212	0	0	0	1
Circa (1950s)	6,830	0.052	0.222	0	0	0	1
Circa (1960s)	6,830	0.070	0.256	0	0	0	1
Circa (1970s)	6,830	0.099	0.299	0	0	0	1
Circa (1980s)	6,830	0.078	0.268	0	0	0	1
Circa (1990s)	6,830	0.198	0.398	0	0	0	1
Circa (2000s)	6,830	0.326	0.469	0	0	1	1
Circa (2010s)	6,830	0.130	0.337	0	0	0	1
Movement (Automatic)	6,830	0.544	0.498	0	0	1	1
Movement (Mechanical)	6,830	0.334	0.472	0	0	1	1
Movement (Quartz)	6,830	0.122	0.327	0	0	0	1
Watch Dimensions (in mm)	6,830	36.078	6.904	9	33	40	62
Material (Includes Ceramic)	6,830	0.018	0.133	0	0	0	1
Material (Includes Diamond)	6,830	0.204	0.403	0	0	0	1
Material (Includes Gold)	6,830	0.649	0.477	0	0	1	1
Material (Includes Platinum)	6,830	0.070	0.255	0	0	0	1
Material (Includes Steel)	6,830	0.347	0.476	0	0	1	1
Material (Includes Titanium)	6,830	0.041	0.199	0	0	0	1
Hammer Price (in $1000s)	6,830	21.557	52.769	0.364	3.635	14.815	950.196

Figure 3 Structured Attributes

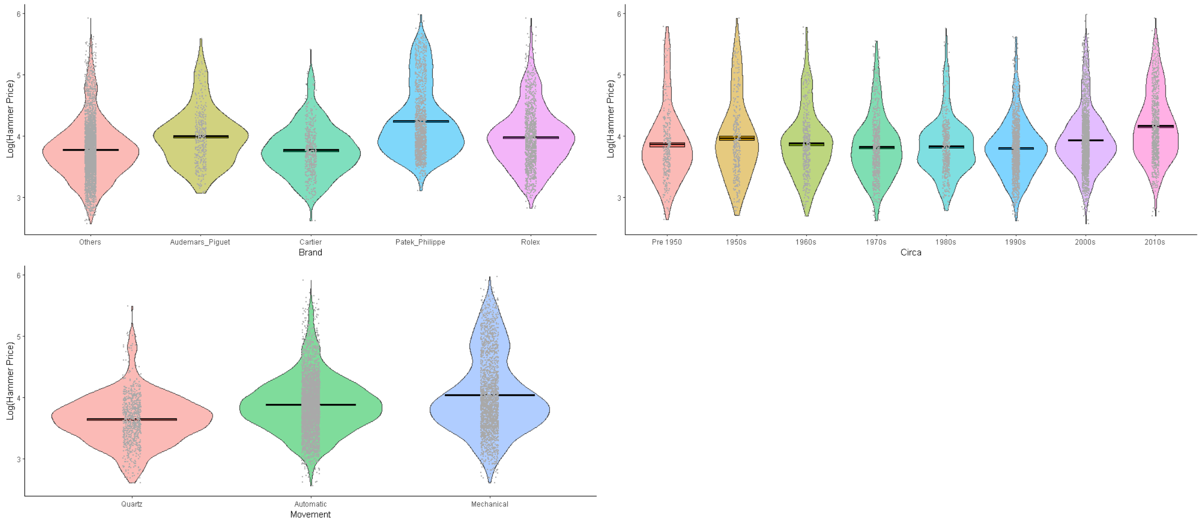


Figure 4 Auction Attributes

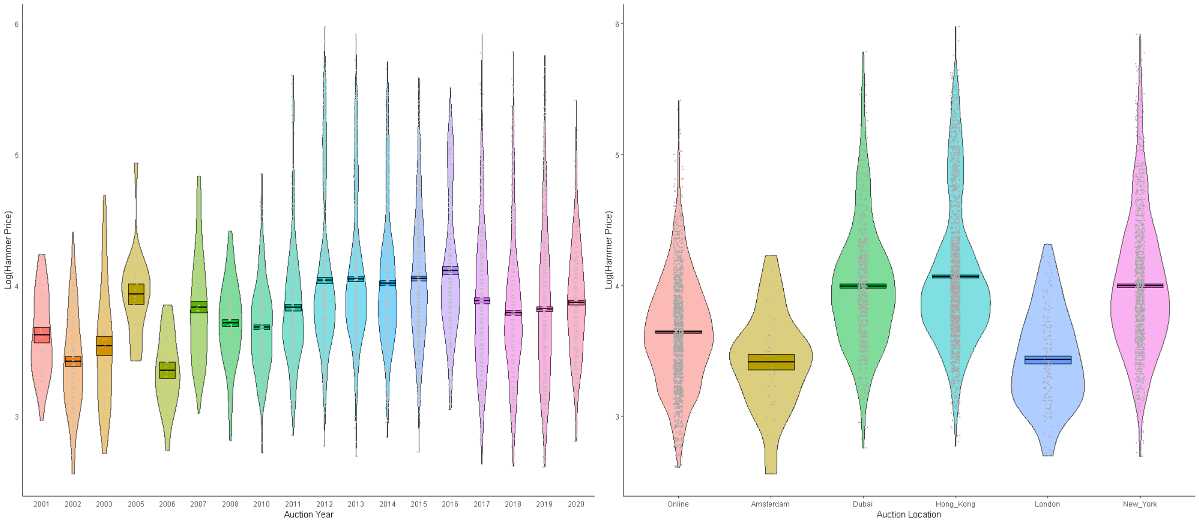
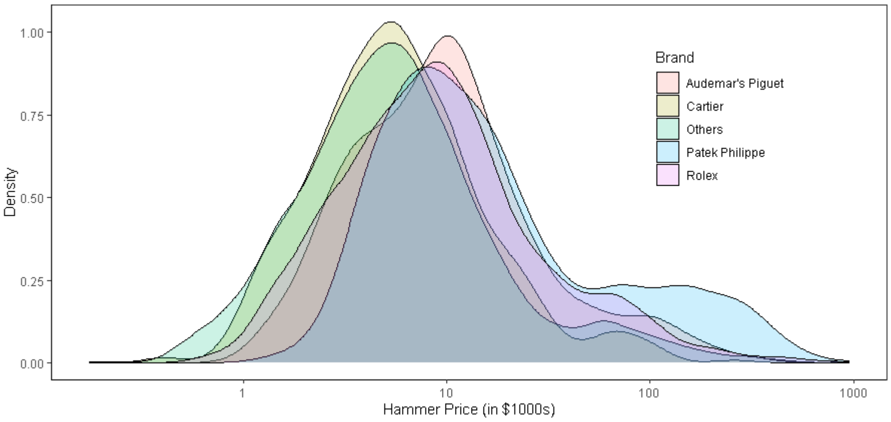


Figure 5 Density Plot of Hammer Price



4.2. Experimental Procedure

We divided the full data into train and test dataset at a ratio of 85:15. We next preprocessed the data by scaling down the image resolutions to 64 by 64 pixels, changing their background to white, and ensuring that all watches within images were vertical in orientation and center in alignment (refer to Figure 2).

We next used the preprocessed training data to train the model. This involved feeding the model the watch images as the unstructured data, hammer price as the economic primitive, structured watch attributes, and structured auction attributes of the train dataset. For the model itself, we specified the number of visual attributes to be discovered as $J = 10$, and tuned the model's hyperparameters; namely, α, β, γ, and δ from Equation 1 along with the number of epochs, batch size, and the learning rate. The hyperparameters of the best model were $\alpha = 1$, $\beta = 6$, $\gamma = 1$ and $\delta = 1000$, epochs=200, batch size=64 and learning rate=0.002. These hyperparameters were obtained by both quantitative loss metrics as well as visual inspection of the ten visual attributes of watches in the training dataset. We provide general intuition on setting a low or a high batch size, epochs and learning rate. A low batch size increases the time required to train the model till convergence while a large batch size significantly degrades the quality of the model so that it is not generalizable beyond the training dataset. Training for low number of epochs may result in the model not converging while training for a very high number of epochs may result in the model overfitting on the train dataset. A very low learning rate can lead the model to get stuck on a local minima or converge very slowly and a very high learning rate can lead the model to overshoot the minima.

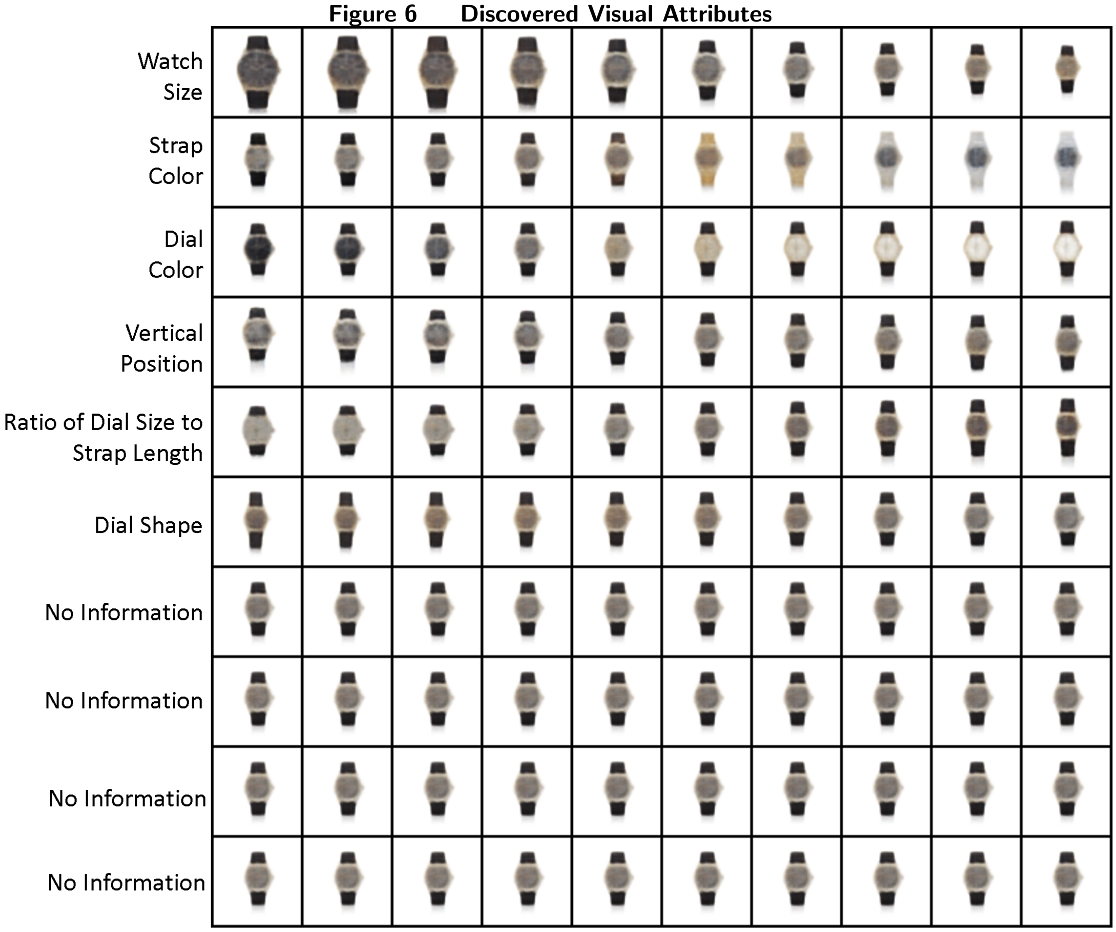
With the now trained model, we used the test dataset to obtain results. Specifically, we "froze" the weights of the neural network and fed the model the testing data; namely, the watch images as the unstructured data, hammer price as the economic primitive, structured watch attributes and structured auction attributes. This allowed us to "discover" visual attributes as well as their values in the test dataset. These visual attributes formed the basis on the contribution of out method, as they are they are *a priori* unknown yet impact economic primitives. In our case, these visual attributes, along with other structured attributes such as brand, were used within a regression to estimate the auctioned watches' hammer prices (i.e., willingness to pay).

4.3. Results

We evaluate the proposed method along three criteria: (1) how well the "discovered" visual attributes can be semantically interpreted, (2) how well the discovered visual attributes are statistically independent, and (3) how well the "discovered" visual attributes help explain the hammer price of the watches when compared to benchmarks that include the known structured watch attributes and structured auction attributes.

4.3.1. Interpretation of Discovered Attributes We evaluate the proposed method along how well humans can semantically interpret "discovered" visual attributes. Specifically, we use the model to generate images showing how changing one visual attribute while keeping others fixed visually changes the image. Figure 6 shows an example output

Figure 6 Discovered Visual Attributes



of "discovered" visual attributes. In each row of Figure 6, we use the model to output changes in attribute values of one visual attribute while keeping the others fixed.

The first visual attribute corresponding to the first row can be semantically interpreted as the size of the watch image. We note that the size of the watch image could be due to both due to the actual size of the watch as well as the distance from which the photograph was taken. The second visual attribute corresponding to the second row can be semantically interpreted as the strap color. The third visual attribute corresponding to the third row can be semantically interpreted as the dial color. The fourth visual attribute corresponding to the fourth row can be semantically interpreted as vertical position. The fifth visual attribute corresponding to the fifth row can be semantically interpreted as the ratio of the dial size to strap length. The sixth visual attribute corresponding to the sixth row can be semantically interpreted as the dial shape. Note there is no visual change in the last four rows. This is due to the model discovering six visual attributes to describe the watch

Figure 7 Correlation Heatmap of "Discovered" Attributes

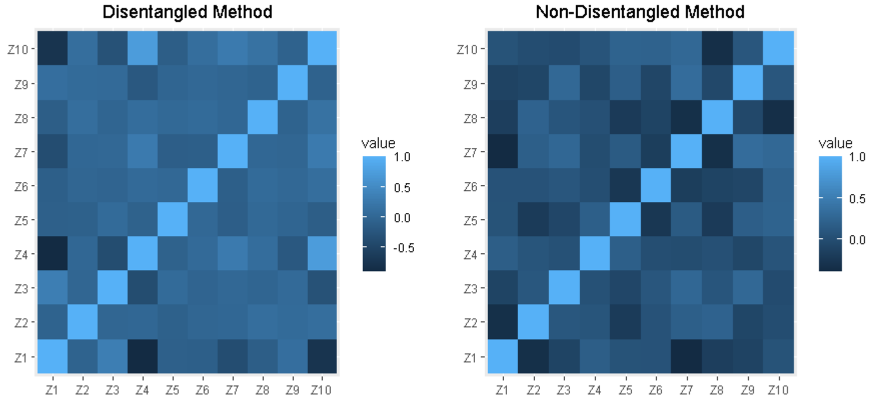
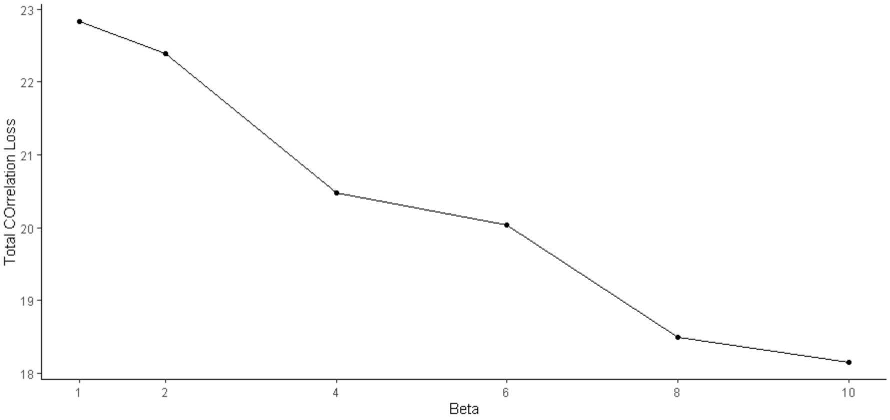


Figure 8 Total Correlation Loss v Beta



images while the rest don't provide any meaningful information. Table 2 has the summary statistics of the visual attribute levels.

Table 2 Summary Statistics of Discovered Visual Attributes (Train + Test Data)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Watch Size	6,830	0.820	0.788	−1.825	0.267	1.359	3.480
Strap Color	6,830	0.255	0.952	−3.998	−0.354	0.850	4.518
Dial Color	6,830	0.353	1.104	−4.275	−0.403	1.112	3.804
Vertical Position	6,830	0.190	0.923	−4.010	−0.403	0.816	3.384
Ratio of Dial Size to Strap Length	6,830	1.115	0.849	−1.919	0.533	1.722	3.971
Dial Shape	6,830	−0.186	0.953	−3.965	−0.830	0.398	4.235

4.3.2. Statistical Independence of Discovered Attributes We next evaluate how well the "discovered" visual attributes are statistically independent both qualitatively and quantitatively. For the qualitative analysis, we plot heatmaps of the correlation matrices of the discovered attributes. If the ten visual attributes discovered from our model are independent, then the correlation matrix of the mean visual attribute values should be close to the identity matrix. Figure 7 shows the heatmap of the correlation matrix with and without disentanglement.

For the quantitative analysis, we use the total correlation between attributes as a measure of their statistical independence. Recall that this measure includes all statistical moments between attributes, not just their linear correlation as the name 'total correlation' may imply. Specifically, it is calculated as the KL-divergence between the joint density of attributes and a the factorized density of attributes in which all attributes are independent of each other. Figure 8 shows the value of the total correlation of "discovered" visual attributes as we change the parameter β. Higher values of β lead to more statistical independence, while lower values lead to more dependence.

4.3.3. Prediction of Economic Primitive (Willingness to Pay) We use a simple linear model to predict the hammer price y_i of the watch for each auction i. The model is specified in Equation (2) with parameter vector $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$, the econometric parameters estimated using Equation (1).

$$\log(y_i) = \mathbf{w}_1^T \mathbf{v}_i + \mathbf{w}_2^T \mathbf{z}_i + \mathbf{w}_3^T \mathbf{a}_i + \epsilon_i \tag{2}$$

where y_i is the hammer price of the auctioned watch expressed in year 2000 dollars; $\mathbf{v}_i$ are the structured product attributes, namely brand of the watch, circa of the watch, movement of the watch, dimensions of the watch and the the material of the watch; $\mathbf{z}_i$ are the visual attributes, namely size of the watch image, strap color, dial color, vertical position, ratio of the dial size to strap length, dial shape and four no information attributes; and $\mathbf{a}_i$ are the auction attributes, namely auction location and auction year; and ϵ is the error term.

We specify a series of nested models based on the linear specification. In the first model (1), we only include structured product attributes. In the second model (2), we only include visual attributes. In the third model (3), we include structured product attributes and visual attributes. In the fourth model (4), we include structured product attributes, visual attributes and auction attributes. We run these four models on the training dataset to recover the coefficients of each attribute.

The results are detailed in Table 3. From model (1), we can see that structured product attributes are only able to explain 38% of the variation in the hammer prices (as characterized by R^2). From the results of model (2), we see that visual attributes are able to explain 58% of the variation in the hammer prices, which is greater than the variance explained by structured attributes only. This is noteworthy because in most marketing models, researchers use only structured attributes and in models based on Berry (1994), Berry et al. (1995), they aggregate across all the visual (and other) attributes to form one attribute that is the unobserved product quality. This leads to two issues. First is the obvious problem of aggregation, which implies the researcher cannot semantically interpret or identify the source of this term. Second, since such an attribute is treated as a structural error, we cannot model consumer heterogeneity across the attribute. In contrast, in our model these unobserved visual attributes are recovered as part of our algorithm, and have semantic interpretation.

Next, we include both structured product attributes and discovered visual attributes in model (3). We observe that structured product attributes and visual attributes together are able to explain $R^2 = 67\%$ of the variation in hammer prices. Finally, in model (4), we include the attributes of the auction as well. We can see that structured product attributes, visual attributes and auction attributes together are able to explain $R^2 = 68\%$ of the variation in the hammer prices. These results taken together suggest that structured product attributes and visual attributes taken together are able to explain the variation in the hammer prices much better than either of the attributes alone.

Our model results show that watches with brand Patek Philippe have the highest hammer prices, followed by Rolex, Audemar's Piguet, and then Cartier compared with the baseline of other brands. The year when the watch was manufactured or circa don't explain the hammer prices. Mechanical watches fetch higher hammer prices than automatic watches. Both mechanical and automatic watches fetch higher hammer prices than the baseline of quartz watches. A larger watch dimension (in the structured data) leads to a higher hammer price. Evaluating the watch material, we find that platinum watches have the highest hammer prices, followed by diamond, gold, titanium, ceramic and finally steel. Much of these results are consistent with prior knowledge about the market for high-end watches.

We find that all of the 6 visual attributes discovered by our approach have an impact on Willingness to Pay that is statistically significant. The attributes and summary statistics

Table 3 Results

	Dependent variable:			
	Willingness to Pay (Highest Bid in Auction)			
	(1)	(2)	(3)	(4)
Constant	2.364*** (0.051)	4.092*** (0.013)	3.514*** (0.045)	3.452*** (0.080)
Brand (Audemar's Piguet)	0.295*** (0.024)		0.163*** (0.018)	0.158*** (0.018)
Brand (Cartier)	0.117*** (0.023)		0.135*** (0.017)	0.131*** (0.017)
Brand (Patek Philippe)	0.537*** (0.016)		0.315*** (0.012)	0.314*** (0.012)
Brand (Rolex)	0.450*** (0.017)		0.225*** (0.013)	0.227*** (0.013)
Circa (1950s)	0.090** (0.036)		0.026 (0.026)	0.027 (0.026)
Circa (1960s)	0.093*** (0.033)		0.020 (0.025)	0.013 (0.025)
Circa (1970s)	0.030 (0.032)		−0.054** (0.024)	−0.056** (0.024)
Circa (1980s)	0.039 (0.033)		−0.041* (0.025)	−0.044* (0.025)
Circa (1990s)	0.115*** (0.030)		−0.003 (0.022)	0.003 (0.022)
Circa (2000s)	0.174*** (0.030)		−0.004 (0.022)	−0.002 (0.022)
Circa (2010s)	0.341*** (0.032)		0.075*** (0.024)	0.071*** (0.024)
Automatic Movement	0.019 (0.021)		0.045*** (0.015)	0.040*** (0.015)
Mechanical Movement	0.237*** (0.021)		0.142*** (0.016)	0.140*** (0.016)
Watch Dimensions	0.028*** (0.001)		0.009*** (0.001)	0.007*** (0.001)
Has Steel	−0.130*** (0.019)		−0.114*** (0.014)	−0.104*** (0.014)
Has Gold	0.109*** (0.019)		0.097*** (0.014)	0.105*** (0.014)
Has Diamond	0.226*** (0.016)		0.155*** (0.012)	0.145*** (0.012)
Has Platinum	0.437*** (0.025)		0.196*** (0.019)	0.205*** (0.019)
Has Titanium	0.125*** (0.031)		0.039* (0.023)	0.037 (0.023)
Has Ceramic	−0.070 (0.044)		−0.057* (0.032)	−0.066** (0.032)
Visual Attribute 1		−0.427*** (0.007)	−0.361*** (0.007)	−0.349*** (0.007)
Visual Attribute 2		0.100*** (0.005)	0.082*** (0.005)	0.081*** (0.005)
Visual Attribute 3		0.173*** (0.005)	0.161*** (0.005)	0.153*** (0.005)
Visual Attribute 4		0.118*** (0.006)	0.073*** (0.005)	0.065*** (0.006)
Visual Attribute 5		−0.070*** (0.006)	−0.052*** (0.006)	−0.047*** (0.006)
Visual Attribute 6		−0.072*** (0.005)	−0.069*** (0.005)	−0.070*** (0.006)
Visual Attribute 7		−0.244*** (0.005)	−0.216*** (0.005)	−0.211*** (0.006)
Visual Attribute 8		−0.071*** (0.004)	−0.055*** (0.004)	−0.030*** (0.005)
Visual Attribute 9		0.019*** (0.005)	−0.012*** (0.005)	−0.011** (0.005)
Visual Attribute 10		0.133*** (0.005)	0.091*** (0.005)	0.078*** (0.006)
Auction Location (Amsterdam)				0.082 (0.061)
Auction Location (Dubai)				0.136*** (0.021)
Auction Location (Hong Kong)				0.150*** (0.025)
Auction Location (London)				−0.050 (0.064)
Auction Location (New York)				0.144*** (0.022)
Auction Year 2002				−0.026 (0.091)
Auction Year 2003				0.123 (0.091)
Auction Year 2005				0.082 (0.098)
Auction Year 2006				−0.064 (0.108)
Auction Year 2007				0.064 (0.072)
Auction Year 2008				0.022 (0.071)
Auction Year 2010				−0.096 (0.067)
Auction Year 2011				−0.038 (0.067)
Auction Year 2012				0.009 (0.066)
Auction Year 2013				0.009 (0.065)
Auction Year 2014				0.053 (0.066)
Auction Year 2015				−0.011 (0.066)
Auction Year 2016				−0.003 (0.067)
Auction Year 2017				0.059 (0.067)
Auction Year 2018				0.044 (0.068)
Auction Year 2019				0.023 (0.067)
Auction Year 2020				0.119* (0.070)
Observations	5,803	5,803	5,803	5,803
R^2	0.382	0.576	0.672	0.679
Adjusted R^2	0.380	0.575	0.670	0.676
Residual Std. Error	0.420 (df = 5782)	0.347 (df = 5792)	0.306 (df = 5772)	0.303 (df = 5750)
F Statistic	178.521*** (df = 20; 5782)	786.895*** (df = 10; 5792)	393.528*** (df = 30; 5772)	234.320*** (df = 52; 5750)

Note: See Figure 6 for Visual Attributes $1, 2, \ldots, 10$. *p<0.1; **p<0.05; ***p<0.01

of their distributions are detailed in Table 2. Connecting this data to Figure 6, we note that for each row, moving from left to right is equivalent to moving in latent space from the lowest (most negative) value of the attribute to the highest (most positive). Thus, since Visual attribute 1 (watch size) has a negative coefficient in Table 3, and in Figure 6 (Row 1), the watches decrease in size as we move from left to right, we infer that larger sizes of watches in the images is associated with higher willingness to pay. Similarly, we find the coefficient for strap color to be positive. As the strap color moves from black to silver to gold, we find that is associated with greater willingness to pay. Another interesting visual attribute is the ratio of dial size to strap length, which we find is decreasing as we move from left to right, and we find that willingness to pay is higher when the dial is larger relative to the strap.

With regard to the location of watches, it is important to control for this since the market demand might vary significantly across location. We find that watches auctioned in Hong Kong have the highest unconditional hammer prices, followed by New York and then Dubai, with Amsterdam and London not different from the baseline of an online auction. The year when the auction takes place doesn't explain any variation in the hammer prices as they are all not statistically significant (excluded from the table).

We then predict the hammer price of the watches in the test dataset using these four models. Table 4 shows the R^2, $RMSE$ (Root Mean Squared Error), RSE (Residual Standard Error), MAE (Mean Absolute Error), AIC (Akaike's Information Criteria) and BIC (Bayesian Information Criteria) of these four models. R^2, $RMSE$ and MAE are calculated using the test dataset. RSE, AIC and BIC are calculated using the train dataset.

There are a few observations worth noting here. First, we observe that in the training data, having only visual attributes (Model 2) had explained a lower variance than only structured attributes (Model 1). However, in the prediction of WTP using the test data, the structured characteristics actually explained more variance. This could well be because there is far more variability in the discovered visual attributes, and a watch model that is present only in the test data might not be as well approximated visually by all the watches present in the training data.

Second, we observe that the result that adding both visual and structured attributes improves the variance explained, since Model 3 obtains a higher R^2 than Model 1 or 2. Third, we observe that the observation above holds not matter which metric we examine, indicating its generality and robustness. Finally, while adding auction variables improves almost all the metrics, it actually results in an increase in BIC, indicating that the additional model complexity outweighs the improved performance using that metric.

Table 4 Prediction

Model	R^2	RMSE	RSE	MAE	AIC	BIC
Model 1	0.366	0.425	0.420	0.323	6410.455	6557.11
Model 2	0.231	0.488	0.347	0.363	4201.685	4281.678
Model 3	0.393	0.423	0.306	0.313	2758.691	2972.007
Model 4	0.408	0.417	0.303	0.309	2663.923	3023.894

5. Discussion and Conclusion

We detail a new approach to discovering unknown product attributes from unstructured data. By unknown, we refer to the fact that both the actual attributes and their number are not specified by the researcher, but actually automatically obtained from our algorithm. There is a large and fast-growing amount of unstructured data, and obtaining attributes from them has the potentially to discover unknown aspects of consumer preferences, and also quantify the strength of their preferences in terms of economically relevant variables. Most existing marketing research involving unstructured data uses machine learning algorithms to obtain a new variable that is typically specified by the researcher, whereas our approach obviates the need for such judgment, which can lead to biased inference.

Our primary contribution is methodological. From a marketing viewpoint, we introduce a method that is able to discover product attributes that impact consumer preferences and economic primitives such as willingness to pay. The machine learning literature on disentanglement has faced an substantial obstacle in the form on impossibility theorems that have shown that using only unstructured data, there are an infinite set of possible generative models. To overcome this challenge, we introduce economically-meaningful supervision to otherwise unsupervised disentanglement methods, thereby overcoming these impossibility theorems.

Our approach is aimed at being immediately practical and useful to managers across a wide set of business domains, especially in marketing. It can take a collection of unstructured data and be able to automatically discover relevant product or service attributes and connect them to economically important variables. Our disentanglement method also obtains attributes that are independent of each other and also semantically meaningful an interpretable, whereas prior work has usually required human intervention to do so. The advantage of our integrated machine learning and econometrics modeling approach is that we also obtain the impact of the discovered attributes on the economic variable. In addition, our model allows the researcher to conduct counterfactuals by examining the economic impact of a change in the design of discovered attributes.

Our approach has a number of limitations. First, it requires an economically relevant variable like demand or Willingness to Pay, otherwise it just reduces to a standard disentanglement problem. However, obtaining reliably truthful observations to economically-relevant variables is often fraught with selection issues and issues with missing observables in real-world contexts. Indeed, a primary reason for our choice of watch auctions was the format of relatively non-substitutable products with hammer prices. Second, the researcher would need to collect the unstructured data in a way that matches the economic variable of interest. Third, there are some initial steps to pre-process the unstructured data so that it is amenable to deep learning. Finally, we might need a lot of unstructured data, although in our application, we were able to obtain reasonable results with a few thousand images, which would be practical for most firms or researchers.

There are a number of potentially avenues for future work. First, given the relative generality of the proposed method, we note that its application to textual data may be more than possible, but actually the data modality that is most fruitful in generating managerial insight. Several recent methods using textual user-generated data have advanced marketing science in areas such as identifying consumer needs (Timoshenko and Hauser 2017) and inferring latent structure of consumer ratings (Chakraborty et al. 2019). These methods, however, have required some degree of human intervention, which perhaps may be ameliorated or complemented by our approach. Second, we note that the particular conditional independence structure assumed in our method may be restrictive (Maaløe et al.. Investigating alternate conditional independence assumptions between structured and unstructured attributes, and their associated modeling requirements, may prove to be an interesting direction for future study. Finally, and perhaps most importantly, we may be able to apply these methods in a manner that is useful to marketers and managers. For example, could one show evidence that the method can discover attributes that are near impossible to discover from only human expertise and intuition, yet are critical for understanding and predicting consumer preferences? It is this in this last direction that our approach is ultimately motivated by–the manager aiming to improve her understanding of the consumer.

References

Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35(8):1798–1828.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

Berry ST (1994) Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* 242–262.

Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* 112(518):859–877, ISSN 0162-1459, 1537-274X, URL http://dx.doi.org/10.1080/01621459.2017.1285773, arXiv: 1601.00670.

Burgess C, Higgins I, Pal A, Matthey L, Watters N, Desjardins G, Lerchner A (2018) Understanding disentangling in β-vae. *arXiv: Machine Learning* .

Burnap A, Hauser JR, Timoshenko A (2019) Design and Evaluation of Product Aesthetics: A Human-Machine Hybrid Approach (ID 3421771), URL https://papers.ssrn.com/abstract=3421771.

Chakraborty I, Kim M, Sudhir K (2019) Attribute sentiment scoring with online text reviews: Accounting for language structure and attribute self-selection Available at SSRN: https://ssrn.com/abstract=3395012 or http://dx.doi.org/10.2139/ssrn.3395012.

Chen RT, Li X, Grosse RB, Duvenaud DK (2018) Isolating sources of disentanglement in variational autoencoders. *Advances in Neural Information Processing Systems*, 2610–2620.

Cho H, Hasija S, Sosa M (2015) How Important is Design for the Automobile Value Chain? SSRN Scholarly Paper ID 2683913, Social Science Research Network, Rochester, NY, URL https://papers.ssrn.com/abstract=2683913.

Dotson JP, Beltramo MA, Feit EM, Smith RC (2019) Modeling the effect of images on product choices. *Available at SSRN 2282570* .

Green PE, Krieger AM, Wind Y (2001) Thirty years of conjoint analysis: Reflections and prospects. *Interfaces* 31(3_supplement):S56–S73.

Guadagni PM, Little JD (1983) A logit model of brand choice calibrated on scanner data. *Marketing science* 2(3):203–238.

Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, Lerchner A (2017) beta-vae: Learning basic visual concepts with a constrained variational framework. *ICLR*.

Kang N, Ren Y, Feinberg F, Papalambros P (2019) Form function: Optimizing aesthetic product design via adaptive, geometrized preference elicitation. *arXiv preprint arXiv:1912.05047* .

Kim H, Mnih A (2018) Disentangling by factorising. *International Conference on Machine Learning*, 2649–2658 (PMLR).

Kingma DP, Welling M (2014) Auto-encoding variational bayes. *stat* 1050:1.

Locatello F, Bauer S, Lucic M, Raetsch G, Gelly S, Schölkopf B, Bachem O (2019) Challenging common assumptions in the unsupervised learning of disentangled representations. *international conference on machine learning*, 4114–4124 (PMLR).

Maaløe L, Sønderby CK, Sønderby SK, Winther O (????) Auxiliary Deep Generative Models. *arXiv:1602.05473 [cs, stat]* URL http://arxiv.org/abs/1602.05473.

Roweis S, Ghahramani Z (1999) A unifying review of linear Gaussian models. *Neural computation* 11(2):305–345, URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6790691, 00715.

Timoshenko A, Hauser JR (2017) Identifying Customer Needs from User-Generated Content. *SSRN Electronic Journal* ISSN 1556-5068, URL http://dx.doi.org/10.2139/ssrn.2985759.