

Think out of the package: Recommending package types for e-commerce shipments

Karthik S. Gurumoorthy^{✉1}, Subhajit Sanyal¹, and Vineet Chaoji¹

India Machine Learning, Amazon, Bangalore, India
{gurumoor, subhajs, vchaoji}@amazon.com

Abstract. Manifold of product attributes such as dimensions, weight, fragility and liquid content determine the package type used by e-commerce companies to ship products. Sub-optimal package types lead to damaged shipments, incurring huge damage related costs and adversely impacting the company’s reputation for safe delivery. Items can be shipped in more protective packages to reduce damage costs, however this increases the shipment costs due to expensive packaging and higher transportation costs. In this work, we propose a multi-stage approach that trades-off between shipment and damage costs for each product, and accurately assigns the optimal package type using a scalable, computationally efficient linear time algorithm. A simple binary search algorithm is presented to determine the hyper-parameter that balances between the shipment and damage costs. Our approach when applied to choosing package type for Amazon shipments, leads to significant cost savings of tens of millions of dollars in emerging marketplaces, by decreasing both the overall shipment cost and the number of in-transit damages. Our algorithm is live and deployed in the production system, where package types for more than 130,000 products have been modified based on the model’s recommendation, realizing a reduction in damage rate of 24%. Overall, the proposed approach has also helped reduce the carbon footprint.

Keywords: Package type selection · Ordinal enforcement · Discrete optimization · Constrained-Unconstrained formulation equivalence · Trade off parameter selection

1 Introduction

E-commerce companies like Amazon uses several different package types to ship products from warehouses to the customer’s doorstep. These package types vary in the extent of protection offered to the product during transit. Generally, robust package types that provide more protection to the product, resulting in reduced number of package related damages, cost more at the time of shipping due to high material and transportation costs, and vice versa. For instance as shown in Fig. 1, Amazon has the following different package type options listed in increasing order of protection afforded to the product: (i) No Additional Packaging (NAP), (ii) Polybags: polythene bags small (PS) and special (PL), (iii) Jiffy mailer (JM), (iv) Custom pack (CP), (v) Corrugated T-folder box (T), (vi) Corrugated box with

variable height (V), (vii) Corrugated carton box (C). Each package type comes in multiple sizes like small, medium, large and extra-large. The combination of packaging type and size is assigned a barcode, e.g. PS6 to PS9 for small polybags. When an item is ready to be packed, the chosen packaging material and size are used to ship the product.



Fig. 1. Different package types

Packaging related damages can happen during transit or handling by an associate during shipment from the warehouse to the customer. As damages result in degraded customer experience, an extra compensatory amount is often paid to the customer over and above the product's price. As these damaged products need to be sent back to the warehouse, there is an additional return shipment cost. Such damages adversely affect the customer relationship since the company's reputation for reliable delivery is impacted. For instance, customers who are dissatisfied with time-critical purchases (e.g., during festivities), may hesitate to buy products in the future. Even the company's relationship with the sellers get impacted, particularly if new products suffer repeated damages over multiple shipments, since the first few customer experiences are critical to the seller's long term success on the platform. The sum of all these costs associated with damages including the product price, reshipment cost, customer compensation cost etc. will henceforth be referred to as damage cost C_{damage} .

To reduce the cost of damages, items can be packed in more protective packaging. However, more protective packaging (e.g. corrugated box (C)) costs more in terms of packaging materials and transportation costs, which could increase the shipping cost that customers have to pay, or costs that the company bears in case of free shipping. It also generates packaging waste at the customer's end which needs to be disposed off additionally. Hence the problem that needs to be addressed is: *"What is the right package type say, between polybag, jiffy mailer or different corrugated boxes that should be used for shipping a product with the best trade off between shipping cost and damage cost ?"* Once the package type is chosen, the smallest container (size) of that package type that could fit the product snugly would be used for actual shipment. This reduces the shipment volume and hence the shipping cost C_{ship} .

1.1 Contributions

Below, we list our main contributions in this work:

(i) We propose a two-stage approach to recommend the correct package type for products resulting in significant savings, primarily from decreased packaging related transit damages. Interestingly, our model’s recommendation also leads to decreased shipping cost compared to the current selection of package types, the reason for which is explained in Section 4.2.

(ii) Our framework naturally provides a scalable mechanism for the package type recommendation, circumventing manual intervention at every stage and deprecating the existing keyword based approach of mapping package type explained in Section 2.1, which is slow, reactive and often subjective.

(iii) We establish novel theoretical connections between the constrained (Ivanov) and unconstrained (Tikhonov) formulations for our unique setting where the optimization variable is discrete, and show that while the constrained formulation is NP -complete, the unconstrained formulation enjoys a linear time solution. Though such connections based on the Lagrange dual formulation are known when the optimization variable is continuous [13], the proof methodology employed in our work to derive similar equivalences when the optimization variable is discrete requires fundamentally new insights into the solution space. To the best of our knowledge, this connection is unknown and not exploited before.

(iv) Our understanding of the solution space further enables us to consistently choose the hyper-parameter for the unconstrained formulation using a simple binary search type algorithm, which optimally provides the trade-off between the different cost parameters.

To summarize, we provide a scalable approach for choosing the best package type for products and also present an efficient algorithm to select the hyper-parameter involved in the optimization.

2 Related work

2.1 Existing packaging selection process

The decision to choose the package type for a product is currently based on a Keyword Based Approach (KBA), where the products are mapped to package types based on whether their title contains a predefined set of positive and negative keywords. Positive keywords work as enablers to ship products in inferior packaging types like polybags or NAP. Examples for positive keywords are helmet, diapers, mosquito net, bag pack, laptop sleeve, bedsheet, cushion, etc. with the assumption being that such products will have low in-transit damages due to inferior packaging. Negative keywords on the other hand prohibit opting for polybag. Examples for negative keywords are bone-china, detergent, harpic, protein supplements, etc. After manually analyzing the product titles, a suitable package type is identified. Another approach that is closely followed is the selection of package types using the historical data on damages. Here, the damage rates of products are collected on a monthly basis and based on set guardrails,

the packaging rules are modified for products with high damages. In addition to being a slow, manual process, this is a reactive approach which does not work for many new products or products whose attributes have been modified recently.

2.2 Why not ordinal regression?

As the different packaging options can be graded in terms of their robustness, the package type forms an ordinal variable with implicit relative ordering between them. This observation naturally surfaces the following question: *“Is predicting the optimal package type for a product just an instance of ordinal regression?”*. Though the answer appears to be a *yes*, the problem lies in the lack of training data. The current assignment of package type to a product is known to be sub-optimal for most of the products w.r.t. trade off between shipping and damage costs. The ideal setting would demand that we have enough samples for every $\langle \text{product, package type} \rangle$ pair, so that one could assign true package type as the target label and perform ordinal regression on product features. This model could then be leveraged to predict package type for new products. Such an exercise would incur significant cost especially at the scale at which e-commerce company like Amazon operates and hence is practically infeasible. The lack of such ground truth data precludes us from performing ordinal regression analysis. We allude to this fact in Section 4.2.

2.3 Comparison with standard machine learning approaches for package planning

The work in [8] shows the adoption of machine learning (ML) in the manufacturing industry for automated package planning. Given a training data with well defined labels of package type to be used for product parts, the goal in these applications is to train a supervised ML model based on product characteristics, which are later used to predict package type for unseen products. Our current work differs from these approaches on the following factors:

- (i) As described above, we do not have any ground truth data to learn a supervised ML model that directly predicts the package type given the product features. We have training label only at the shipment level that informs whether a product shipped in particular package type was damaged (1) or not (0).
- (ii) There is natural ordering between the package types that should be enforced in any learning algorithm.

Given the above two constraints, we are not aware of any learning based framework that automatically chooses the best package type in linear time and is scalable to millions of products.

A majority of the work in logistics is around space optimization, which is broadly related to bin *packing* algorithms [12], [11], not to be confused with the *packaging* type selection problem. The aim of the former is to identify those set of products, each of a specific volume, that should be loaded together in a container, in a specific orientation, so that the number of container used is minimum. The bin packing problem has no notion of choosing the best package

type for each product. Our work also has very little connection with the box size optimization problem [18], where the goal is to determine the best box sizes that should be used to ship the products, so that the total shipment volume across all product is minimum. We do not optimize for the different sizes of the packages in the current work, but rather determine which package type is best suited for a product. Likewise, we do not forecast packaging demand like the methods developed in [1].

3 Two-stage approach for optimal package selection

3.1 Stage 1: Estimating the transit damage probability of a product given a package type

In this stage, we build a model to solve the following problem, “*Given a product and a package type, what is the probability that a shipment of the product with that package type is likely incur costs due to damages?*”. These damage probabilities are computed for every <product, package type> pair as a product may never have been shipped using a particular (say hitherto unknown optimal) package type to directly retrieve it from the shipment data. In short, our model predicts $p(d|i, j)$ where i refers the product, j refers to the package type and d is a variable indicating damage in transit, with $d = 0$ denoting no damage and $d = 1$ specifying a damage in transit.

For modeling, we considered historical shipment data where for every shipment we have a binary flag a.k.a. the target label indicating whether the shipment resulted in package related damage. We built this model using various metadata associated with the product as predictor variables. The following enumerate a sample set of attributes: product title, category, subcategory, product dimensions, weight, hazardous flag (indicating if product pertains to hazardous materials), fragile flag (denoting whether the product is fragile), liquid flag (representing if the product contains liquids), % air in shipment computed as the difference between the package volume and the product volume, etc. Based on the above set of features we trained a model to predict the probability that the shipment using the particular package type will incur a damage.

Maintaining ordinal relationship between different package types The notion of graded robustness between package types correlates with the cost of the packaging material where the cost of packaging goes up if we opt for a more robust package type and vice versa. As there exists an ordinal relationship among the various package types, i.e. they can be ordered in terms of their associated robustness, we need to impart this notion to our model while estimating the damage probabilities. Let m and n be the number of products and package type respectively and without loss of generality let the package type j_k be inferior to j_{k+1} represented by the ordinal relationship $j_1 \leq j_2 \leq \dots \leq j_n$. During modeling we need to ensure that $p(d|i, j_{k+1}) \leq p(d|i, j_k)$ for all products i . In other words, the prediction function needs to be *rank monotonic* [9] where

the rank denotes the robustness of the package type. Note that we require the predictions to satisfy the ranking relationship only between the different package types associated for a given product and not across two different products. We achieve rank monotonicity by two means: (a) Augmenting the training data, and (b) Proper representation of the package type feature and imposing lower bound constraints on the corresponding model coefficients.

Firstly, we append the modeling data as follows:

(i) For every damaged shipment, we create additional shipments with the same product and other inferior (less robust) package types and consider them to be damaged as well. This is to incorporate the notion that if a shipment of a product gets damaged with a particular package type, it is likely to get damaged in package types which are inferior in terms of robustness.

(ii) Likewise, for every shipment without any packaging related damages, we artificially introduce more shipments with the same product and other superior (more robust) package types and consider them to be not damaged as well. This is to incorporate the notion that if a shipment of a product does not get damaged with a particular package type, it is unlikely to get damaged in superior package types.

Appending the data set has an added advantage of creating many more samples for the positive damaged class (label = 1), as typically very few shipments, less than 0.6%, incur packaging related damages. This in turn reduces the model variance as even the positive class is well represented. Secondly, expressing the damage probability values in terms of the sigmoid function, namely $p(d|i, j) = \frac{1}{1 + \exp(-f(\mathbf{z}_i, j))}$ where \mathbf{z}_i denote the rest of input features barring the package type, we represent $f(\cdot)$ as $f(\mathbf{z}_i, j) = g(\mathbf{z}_i) + \beta_j$. Here $\{\beta_j\}_{j=1}^n$ are the n model coefficients corresponding to each package type. Ensuring rank monotonicity is equivalent to constraining $\beta_k \geq \beta_{k+1}$. Expressing $\beta_k = \beta_{k+1} + \epsilon_k$, we enforce that $\epsilon_k \geq 0, \forall k \in \{1, 2, \dots, n-1\}$. In the event that $g(\cdot)$ is linear, i.e., $g(\mathbf{z}_i) = \mathbf{w}^T \mathbf{z}_i$ as the case with Logistic Regression classifier, then for each package type j_k , we append the feature vector \mathbf{z}_i to create $\tilde{\mathbf{z}}_{ik} = [\mathbf{z}_i, \mathbf{p}_k]$ where $\mathbf{p}_k = \underbrace{[0, 0, \dots, 0]}_{k-1}, \underbrace{[1, \dots, 1]}_{n-k+1}$, augment the model coefficient vector \mathbf{w} to $\tilde{\mathbf{w}} = [\mathbf{w}, \epsilon_1, \dots, \epsilon_{n-1}, \beta_n]$, and express $f(\mathbf{z}_i, j_k) = \tilde{\mathbf{w}}^T \tilde{\mathbf{z}}_{ik}$. The vector $\tilde{\mathbf{w}}$ is determined as part of the model training process under the constraint that $\epsilon_k \geq 0, \forall k$.

3.2 Stage 2: Identifying the optimal package type for each product

Optimally assigning the packaging type for each product involves finding the right balance between adopting a robust packaging and incurring more material and transport costs, and settling for an inferior option with a higher probability of in-transit damages leading to increased damage costs. We formulate this trade-off as an optimization problem. Given a packaging type assignment j for a product i , the packing material cost $m(i, j)$ and the transportation cost $s(i, j)$ are known and readily available. The quantity $s(i, j)$ is known as the *bill weight* and is proportional to the package volume. The net shipping cost,

$C_{ship}(i, j) = m(i, j) + s(i, j)$. The total shipment cost, T_{ship} , computed over all the products equals: $T_{ship} = \sum_i C_{ship}(i, j) * s_{vel}(i)$, where $s_{vel}(i)$ is the sales velocity—number of units sold in a specified period—of the product i . Further, if a product i associated with the package type j is damaged in transit, we incur a net damage cost $C_{damage}(i)$. This damaged cost depends only on the product and independent of the package type used in the shipment. Using the in-transit damage probability $p(d|i, j)$ determined in stage 1 (Section 3.1), we estimate the damage cost as: $T_{damage} = \sum_i p(d|i, j) * s_{vel}(i) * C_{damage}(i)$.

Let us denote the current package type assignment of product i by j^{cur} . According to the current package type assignment, the total cost due to in-transit damages is: $T_{damage}^{cur} = \sum_i p(d|i, j^{cur}) * s_{vel}(i) * C_{damage}(i)$. The objective of the optimization is to determine the package types such that T_{ship} is minimized and at the same time T_{damage} is not largely different from T_{damage}^{cur} i.e., $T_{damage} \leq \gamma * T_{damage}^{cur}$, where $\gamma \geq 0$ sets the allowable tolerance w.r.t. T_{damage}^{cur} and is determined by business requirements.

Mathematical formulation Let the variable x_{ij} indicating whether a product i is to be shipped using the package type j , be the $\langle i, j \rangle$ entry of the binary matrix X . These variables have to satisfy the following constraints, namely: $x_{ij} \in \{0, 1\}$, $\forall i, j$ and $\sum_j x_{ij} = 1$, $\forall i$. The first constraint states that a product is either shipped in a particular type ($x_{ij} = 1$) or not ($x_{ij} = 0$). The second constraint specify that a product should be shipped using one and only one package type. In addition to the aforesaid binary constraints, we also need to specify infeasible conditions that preclude certain products to be shipped via certain modes of packaging. For instance, liquid, fragile and hazardous products can neither be recommended polybags nor be shipped without any packaging if they are not currently shipped in these package options. We enforce these infeasibility constraints by creating a mask matrix M where we set $M_{ij} = 1$ if product i cannot be shipped in package type j and $M_{ij} = 0$ otherwise. By imposing the constraint $\sum_{i,j} M_{ij} * x_{ij} = 0$, the optimization algorithm will be coerced to set $x_{ij} = 0$ whenever $M_{ij} = 1$, thereby meeting our infeasibility requirements. Letting $S_{ij} = C_{ship}(i, j) * s_{vel}(i)$ to be the net shipment cost when product i is sent in package j , $D_{ij} = p(d|i, j) * s_{vel}(i) * C_{damage}(i, j)$ as the net damage cost when the shipment experiences an in-transit damage due to the packaging, $T = \gamma * T_{damage}^{cur}$, our objective can be mathematically expressed as:

$$\min_X \sum_{i,j} S_{ij} * x_{ij} \quad s.t. \quad \sum_{i,j} D_{ij} * x_{ij} \leq T \quad (3.1)$$

where, $x_{ij} \in \{0, 1\}$, $\sum_j x_{ij} = 1$, $\forall i$ and $\sum_{i,j} M_{ij} * x_{ij} = 0$.

However, computing the optimal solution for X based on the Integer Programming (IP) formulation in eq.(3.1) is computationally expensive as it is a known *NP*-complete problem [14]. The IP formulation is definitely not scalable

and is of very limited use for our setting. Hence, we *do not* compute the solution for X by solving eq.(3.1). We present the IP objective with the only intent of mathematically formulating and motivating our optimization problem. The direct minimization of the shipping cost, while enforcing that overall damage cost does not exceed the constant T , makes the setting easier to understand. We abstain from solving for X based on this IP objective.

A closer look into the constraints on X reveals that, the constraints are only intra-product, i.e., across different packaging options for a given product and there are no inter-product constraints. This insight enables us to derive an equivalent formulation for eq.(3.1) whose solution, as we demonstrate, can be obtained via a simple linear time algorithm in $O(mn)$. To this end, let $S(X) = \sum_{i,j} S_{ij} * x_{ij}$, $D(X) = \sum_{i,j} D_{ij} * x_{ij}$, and consider the formulation:

$$\begin{aligned} \min_X E(X) &= S(X) + \lambda D(X) \text{ s.t.}, \\ x_{ij} &\in \{0, 1\}, \sum_j x_{ij} = 1, \forall i \text{ and } \sum_{i,j} M_{ij} * x_{ij} = 0, \end{aligned} \quad (3.2)$$

where the hyper-parameter λ is a single globally specified constant *independent* of the products and the package types. The constrained formulation in eq.(3.1) is known as the Ivanov formulation [6] and the objective in eq.(3.2) is referred to as the Tikhonov formulation [17]. The equivalences between the two are specifically known for Support Vector Machines [13], [3] where the optimization variable, the weight vector \mathbf{w} , is continuous and is based on the Lagrange dual formulation. This approach does not work in our discrete setting where X is binary valued. We need to establish this equivalence without invoking the Lagrange formulation and hence our proof methodology is substantially different.

Let X_λ and X_T be the optimal solutions for the hyper-parameters λ and T in Tikhonov and Ivanov formulations respectively. Under mild conditions on the shipment and damage cost values, we prove that these two formulations are equivalent in the sense that for every T in Ivanov, \exists a value of λ in Tikhonov such that both the formulations have the exact same optimal solution in X . To this end, we have the following lemmas:

Lemma 1. *The value of the objective function $E(X_\lambda)$ at the optimal solution X_λ strictly increases with λ .*

Lemma 2. *The overall damage cost $D(X_\lambda)$ [shipment cost $S(X_\lambda)$] at the optimal solution X_λ is a non-increasing [non-decreasing] function of λ , i.e., if $\lambda_1 \leq \lambda_2$ then $D(X_{\lambda_1}) \geq D(X_{\lambda_2})$ [$S(X_{\lambda_1}) \leq S(X_{\lambda_2})$]. Further, if $X_{\lambda_1} \neq X_{\lambda_2}$ then we get the strict inequality, namely $D(X_{\lambda_1}) > D(X_{\lambda_2})$ [$S(X_{\lambda_1}) < S(X_{\lambda_2})$].*

Lemma 2 states that $D(X_\lambda)$ is a piece-wise constant function of λ whose value decreases when the optimal solution changes. The length of the constant portion equals the range of λ having the same optimal solution. Further, $D(X_\lambda)$ is discontinuous and points of discontinuity occurs at those values of λ for which there are two different optimal solutions in X_λ . Lack of space precludes us from

giving the details of the proof. We establish the equivalence through the following theorems.

Theorem 3. *For every λ in eq.(3.2), $\exists T(\gamma)$ in eq.(3.1) such that $X_\lambda = X_T$.*

We define a quantity Δ to equal the largest change between the two values of $D(X_\lambda)$ at the points of discontinuity. For our specific D matrix, $\Delta \leq \max_i \left[\max_{j, M_{ij}=0} D_{ij} - \min_{j, M_{ij}=0} D_{ij} \right]$. Armed with this definition, we now prove a mildly weaker equivalence in the opposite direction.

Theorem 4. *For every $T = \gamma * T_{damage}^{cur}$ in eq.(3.1) for which the optimal solution X_T exists, one can find a $T^* \in [T, T + \Delta)$ such that for this value of T^* , $\exists \lambda$ in eq.(3.2) satisfying $X_\lambda = X_{T^*}$.*

Linear time algorithm The primary advantage of this equivalence is that the Tikhonov formulation in eq.(3.2) enjoys a linear time algorithm compared to the Ivanov problem in eq.(3.1) which is NP -complete. To see this, note that the constraints in the variables x_{ij} are only across the different package types j given a product i and there are no interaction terms between any two different products. Hence the optimization problem can be decoupled between the products and reduced to finding the optimal solution *independently* for each product agnostic to others. For each product i , define the vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ and consider the optimization problem:

$$\min_{\mathbf{x}_i} \sum_j [S_{ij} + \lambda D_{ij}] x_{ij} \text{ s.t., } x_{ij} \in \{0, 1\}, \sum_j x_{ij} = 1 \text{ and } \sum_j M_{ij} * x_{ij} = 0. \quad (3.3)$$

Among all the package types where $M_{ij} = 0$, the minimum occurs at that value of $j = j_i^*$ where the quantity $S_{ij_i^*} + \lambda D_{ij_i^*}$ takes the least value. In other words, define $j_i^* = \operatorname{argmin}_{j, M_{ij}=0} [S_{ij} + \lambda D_{ij}]$. Then $x_{ij_i^*} = 1$ and $x_{ik} = 0, \forall k \neq j_i^*$ is the optimal solution. As it only involves a search over the n values, its time complexity is $O(n)$ for each product. Hence the optimal solution X_λ across all the m products can be determined in $O(mn)$.

Selection of the hyper-parameter λ It is often easier to specify a bound on the overall damage cost $D(X_\lambda)$ via the tolerance constraint $T = \gamma * T_{damage}^{cur}$ in the Ivanov formulation in eq.(3.1), as it is driven by business requirements such as customer satisfaction, impact of damages on downstream purchase behavior, etc. However, knowledge of γ alone is of little value as the Ivanov formulation being NP -complete, is computationally expensive to solve and we rightly refrain from doing so. Instead, we determine the corresponding λ through an efficient algorithm and then solve the Tikhonov formulation in eq.(3.2) in linear time as explained in Section 3.2. Although no closed form expression exists relating the two, the non-increasing characteristic of $D(X_\lambda)$ in Lemma 2 can be leveraged

Algorithm 1 Algorithm to determine λ given T and stopping criteria ρ

```

function DETERMINELAMBDA( $T, \rho$ )
  Set:  $\lambda_{min} = 0, \lambda_{max} =$  chosen high value,  $\lambda_{mid} = \frac{\lambda_{min} + \lambda_{max}}{2}$ , stoppingCriteria
  = False
  do
    Set:  $\lambda = \lambda_{mid}$ 
    Determine: Optimal solution  $X_\lambda$  using the linear time algorithm.
    if  $D(X_\lambda) < T$  then
       $\lambda_{max} = \lambda_{mid}$ 
    else
       $\lambda_{min} = \lambda_{mid}$ 
    end if
    Recompute:  $\lambda_{mid} = \frac{\lambda_{min} + \lambda_{max}}{2}$ 
    if  $(|\lambda_{mid} - \lambda| \leq \rho)$  or  $(D(X_\lambda) == T)$  then
      Set: stoppingCriteria = True
    end if
  while (stoppingCriteria == False)
  return  $\lambda$ 
end function

```

to design a binary search algorithm for λ , as described in Algo. 1. The crux of our method is to repeatedly bisect the interval for the search space of λ and then choose the subinterval containing the λ . The technique is very similar to the bisection method used to find the roots of continuous functions [4]. The user input ρ in Algo. 1 is the stopping criteria on the minimum required change in λ values between successive iterations for the while loop to be executed. The number of iterations is inversely proportional to the magnitude of ρ .

Package prediction for new products The definition of the net shipping and damage cost matrices includes the sales velocity term $s_{vel}(i)$, as the total shipment and damage costs across all products explicitly depend on the individual quantities of products sold. Hence the optimization problem in eq.(3.1) deliberately makes use of the sales velocity term folded into the S_{ij} and D_{ij} matrix entries. However, for new products, the sales velocity is unknown and needs to be forecasted; which is generally very difficult and at most times noisy [10]. The lack of this term seems to preclude the new products from being part of the optimization in eq.(3.1). However, the equivalent Tikhonov formulation in eq.(3.2) comes to our rescue. Closely looking into the product-wise optimization problem in eq.(3.3), note that $s_{vel}(i)$ appears in the same form (linearly) in both the S_{ij} and D_{ij} quantities and also *does not* depend on the package type j . Hence it can be factored out and dropped from the optimization altogether. The equivalent formulation has revealed a key insight that once λ is appropriately chosen, the optimal solution is *independent* of the sales velocity. Hence for all new products l , we only need to compute quantities $\{S_{lj}, D_{lj}\}_{j=1}^n$ without factoring in $s_{vel}(i)$ and choose that package type j_l^* with the least value of $S_{lj_l^*} + \lambda D_{lj_l^*}$ among the package types where $M_{lj} = 0$.

4 Experimental results

Our training data for stage 1, where we predict the damage probabilities, consists of shipments during a 3 month period in 2019. We augmented the data with artificially induced inferior and superior packaging types and their corresponding 1 and 0 target values. We opted for the Logistic Regression classifier to learn and predict the damage probabilities $p(d|i, j)$, as it enables us to interpret and explain the predictions. Importantly, its linearity (post the link function) endows the model with the notion of ordinal relationship between packages by appending the sample features \mathbf{z}_i with the package related features \mathbf{p}_k as elaborated in Section 3.1. Though our classifier is linear in the feature space, we introduced non-linearity through various polynomial transformations of the input features and having interactions between the product and the package features to create new (non-linear) features. In more than 100 *million* augmented training shipments, only 0.7% shipments belonging to class 1 incurred damages due to packaging related issues. We counter this huge class imbalance by specifying class specific weight values of $1 - \tau$ and τ to classes 1 and 0 respectively to the cross-entropy loss function where we set $\tau = 0.007$. We assessed the performance of our model on a test data consisting of 8 *million* shipments for about 600,000 products, out of which only 0.6% shipments incurred packaging related damages. The shipments in the test data occurred in a different time period w.r.t. training data. After augmenting the test data with artificially induced inferior and superior packaging types, the models performance on the area under the curve (AUC) metric was **0.902**.

4.1 Calibration

Since we are interested in estimating the actual probability of damage rather than binary classification, the estimated raw damage probabilities $p(d|i, j)$ need to be calibrated to reflect the true damage probabilities in the shipment data. This is more so, as we introduced class specific weights during training. We used Isotonic Regression [2] to learn the calibration function. It yielded the smallest average *log-loss* (log-loss = 0.0347) compared to the implicit calibration via the closed form expression derived in [7] (eq.(28)) for binary Logistic Regression models (log-loss = 0.0379), and Platt Scaling [15] (log-loss = 0.0349). The log-loss for each shipment equals: $-y \log(p_{cal}) - (1 - y) \log(1 - p_{cal})$, where y is the actual label and p_{cal} is the calibrated damage probability. All these calibration methods significantly reduces the uncalibrated average log loss of 0.4763. To assess the correctness of post-calibrated values, for each package type we bucketed its shipments into 20 quantiles based on their calibrated values. For each quantile, we computed the absolute difference between the actual damage rate and the average of the calibrated values, weighted these absolute differences proportional to the number of shipments in each quantile, and then summed them. Fig. 2 shows the summed, weighted absolute differences for each package type, for different calibration methods. Observe that Isotonic Regression has the

lowest values across multiple package types. *Such low difference values highlight the estimation accuracy of our post-calibrated damage probability values.*

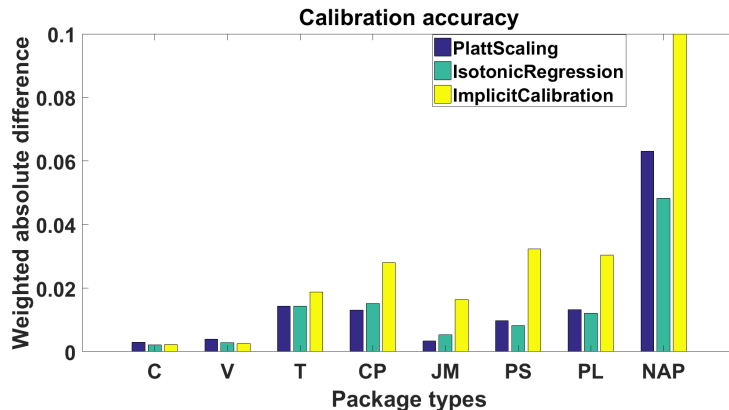


Fig. 2. Weighted absolute difference between estimated and true damage rates

4.2 Package type recommendation

For a dataset of about 250,000 products in more than 10 categories with active purchase history in Amazon, we determined their raw damage probability for all possible package type options and then calibrated them using Isotonic Regression. Table 1 shows the relative average damage probabilities computed across the products for each package type. The damage probability for shipment without packaging (NAP) is set to 1 and the values for other package types are scaled relatively. The business sensitive nature of these damage probabilities precludes us from disclosing their actual estimated values. Observe that our model has indeed learned the implicit ordering between the package types, where the superior package types like C and V have the lowest values and inferior package types like PS and PL have the highest. The predicted damage probabilities are then fed into our optimization algorithm that proposes optimal packaging type for all the products.

For each $\langle \text{product, package type} \rangle$ tuple, we identified the smallest size of that package type that could fit the product snugly. This reduces the shipment volume and also the shipment cost. Recall that by setting entries $M_{ij} = 1$ in the mask matrix M , we can prevent the optimization from choosing the package type j for product i . We set $M_{ij} = 1$ for the following cases based on business rules: (i) products which due to its large size and volume cannot be shipped even in the largest container of certain package types, equivalent to setting the corresponding $S_{ij} = \infty$, (ii) liquid products from being shipped in JM, PS, PL or NAP; restricting fragile products from being sent in T, CP, JM, PS, PL or NAP; disallowing hazardous products to be shipped in PS, PL or NAP if these

products are not currently shipped in these package types (the latter condition is required as these flags can sometimes be erroneously set), (iii) inferior package types compared to the current selection i.e., $j < j^{cur}$ for products (with active purchase history) having high damages in the current package type, (iv) superior package types $j > j^{cur}$ if $S_{ij} > S_{ij^{cur}}$ for products with very low damages in their current packaging type, (v) sensitive products belonging to certain categories from being sent in NAP without any packaging etc.

To corroborate our theoretical results in Lemmas 1 and 2, we ran the Tikhonov formulation in eq.(3.2) for different values of λ , each in linear-time, and plot the results in Fig. 3. The values of $S(X_\lambda)$ and $D(X_\lambda)$ are scaled relative to the total shipment and damage costs from using the current package type, respectively. A value greater (lesser) than 1 indicates that these costs will be higher (lower) compared to the current levels when the products are shipped based on the model recommended package types. Similarly, $E(X_\lambda)$ is scaled relative to the sum of current shipment and damage costs. Observe that the trends of $E(X_\lambda)$, $D(X_\lambda)$ and $S(X_\lambda)$ as we increase λ are in accordance with the claims made in Lemmas 1 and 2. To verify the equivalence relations between the Ivanov and Tikhonov formulations stated in Theorems 3 and 4, we implemented the Integer Programming for Ivanov by setting $\gamma = 1.0$ using the CVXPY package [5]. We then determined the value of corresponding λ by executing our binary search method (Algo. 1) for $\rho = 0.001$ and $\lambda_{max} = 1000$. The algorithm met the stopping criteria in 19 iterations and returned with $\lambda = 0.13387$. The identical results for $(\lambda = 0.13387, \gamma = 1.0)$ in columns VII and VIII of Table 1 is a testimony to this equivalence relationship. We validated this equivalence for other values of γ using our binary search algorithm and obtained similar results. In Fig. 3, note that though the net damage cost $D(X_\lambda)$ for $\lambda = 0.13387$ (and for $\gamma = 1.0$) exactly matches the cost value computed from using the current package types (ratio = 1 marked in horizontal red dotted line), the shipping cost $S(X_\lambda)$ is smaller than the current shipment cost (ratio = 0.843 marked in horizontal green dotted line). In other words, we are able to reduce the shipping cost from the current value without further increasing the damage cost. This again points to the fact that the existing product to package type mappings are sub-optimal, preventing us from pursuing the path of ordinal regression as explained in Section 2.2.

For each package type in Table 1, we show the ratio of number of products mapped to that package type by our algorithm and the number of products currently assigned to the package type, for different λ values. For instance if 100 products are currently shipped in package type C and our model recommends using C for 120 products, the ratio will equal 1.2. A number greater (lesser) than 1 denotes higher (lesser) recommendation of that package type compared to the current usage. Note that as we increase λ giving more importance to damage cost, the ratio for superior package types such as C and V steadily increases, and this trend is reversed for inferior packaging options such as PL and NAP. This shift is as expected since the damage rate and the damage cost decrease at higher λ values. In Table 2 we show the ratio between the number of products

recommended to be sent in a particular package type computed at $\lambda = 1.5$ and the number of products currently shipped in these package types for different product categories. The value 0/0 means no product of that category is currently shipped in the specific package type and our model does not recommend it either. The true counts are confidential and cannot be disclosed. Observe that for liquid, fragile and hazardous products, the ratio is less than 1 for inferior package types such as JM, PS, PL and NAP, indicating that our method recommends lesser usage of these options for these kinds of products. Notably, many electronics products in column VI with high damage probability are moved to the most superior C package type, further contributing to the decreased damage rate of 24% as observed in Section 4.3, and cost savings of tens of millions of dollars even in emerging marketplaces.

Table 1. Relative avg. calibrated damage probabilities and change in product mappings for package types

| I | II | III | IV | V | VI | VII | VIII |
|-----------------------|---------------------|---|---|---|---|---|--|
| | <i>Package type</i> | <i>Relative avg. damage probability</i> | <i>Ratio for $\lambda = 0.5$</i> | <i>Ratio for $\lambda = 1$</i> | <i>Ratio for $\lambda = 1.5$</i> | <i>Ratio for $\lambda = 0.13387$</i> | <i>Ratio for $\gamma = 1$</i> |
| Superior package type | Carton box(C) | 0.022 | 0.915 | 1.176 | 1.314 | 0.499 | 0.499 |
| | Variable height(V) | 0.027 | 0.566 | 0.676 | 0.740 | 0.392 | 0.392 |
| | T-folder(T) | 0.043 | 1.227 | 1.282 | 1.298 | 0.898 | 0.898 |
| | Custom pack(CP) | 0.112 | 2.420 | 2.438 | 2.418 | 2.174 | 2.174 |
| | Jiffy mailer(JM) | 0.174 | 0.408 | 0.503 | 0.539 | 0.236 | 0.236 |
| Inferior package type | Small polybag(PS) | 0.447 | 1.586 | 1.360 | 1.238 | 1.955 | 1.955 |
| | Special polybag(PL) | 0.448 | 1.043 | 0.906 | 0.798 | 0.973 | 0.973 |
| | No packaging(NAP) | 1.0 | 1.748 | 1.144 | 0.940 | 3.635 | 3.635 |

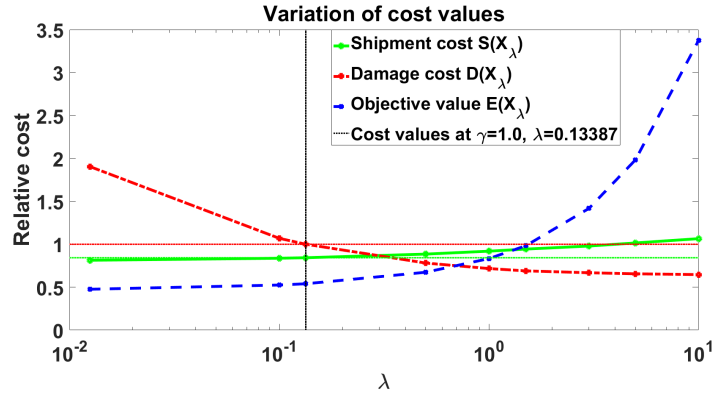


Fig. 3. Variation of relative cost values with λ

4.3 Impact analysis from actual shipment data

The numbers quoted below are excerpts from the actual shipment data, where for 130,000 products contributing to 21% shipments, their current package type was changed to the model's recommendation. We used the proposed package type obtained for $\lambda = 1.5$ ($\gamma = 0.69$), thus giving higher weight to reducing damage costs. The rationale being that receiving damage products negatively affects the customer trust in e-commerce companies and could affect their downstream purchase behavior. When these shipments were compared against those where the original package type was used, we observed the following significant positive impacts: (i) Decrease in damage rate by **24%**, (ii) Decrease in transportation cost per shipment by 5%, (iii) Salability of products undelivered to customer because of transit damages improved by 3.5%. The only negative impact was that the material cost of the shipping supplies increased by 2%, as many products were moved to superior package types to reduce damages.

Table 2. Relative change in product mappings across different categories

| I | II | III | IV | V | VI | VII | VIII |
|-----------------------|-----------------------|------------------------|-------------------------|---------------------------|-----------------------------|-------------------------|------------------------|
| | <i>Package type</i> | <i>Liquid products</i> | <i>Fragile products</i> | <i>Hazardous products</i> | <i>Electronics category</i> | <i>Kitchen category</i> | <i>Beauty category</i> |
| Superior package type | Carton box(C) | 0.929 | 1.794 | 1.975 | 9.525 | 1.214 | 0.927 |
| | Variable height(V) | 0.754 | 1.047 | 0.915 | 1.017 | 0.859 | 0.478 |
| | T-folder(T) | 1.496 | 1.286 | 1.281 | 2.302 | 1.427 | 1.221 |
| ↓ | Custom pack(CP) | 0/0 | 1.383 | 1.081 | 1.873 | 2.855 | 0.818 |
| | Jiffy mailer(JM) | 0.519 | 0.486 | 0.488 | 0.392 | 0.695 | 0.928 |
| | Small polybag(PS) | 0.000 | 0.821 | 0.843 | 1.203 | 1.263 | 1.720 |
| | Special polybag(PL) | 0/0 | 0.541 | 0.500 | 0.863 | 1.124 | 4.000 |
| | Inferior package type | No packaging(NAP) | 0/0 | 0.598 | 0.806 | 0.737 | 0.701 |

5 Conclusion and future work

We presented a two-stage approach to recommend optimal packaging type for products, where we first estimated the calibrated damage probabilities for every <product, package type> tuple and then fed them into our linear-time optimization algorithm to select the best type. The binary search algorithm efficiently computes the trade-off parameter λ given the value γ in the Ivanov formulation.

In many scenarios, the extent of damages depend on the distance shipped, the air/ground mode of transportation used, the quality of roads along the route, the handling by the courier partners, the location of the warehouses or even the time of year as during monsoon seasons, more protection against water or moisture may be needed for some products. In addition, protective packaging could be recommended for specific customers who are highly valued or who had negative delivery experiences in the past. Going forward, we would like to lay emphasis on

predicting the optimal packaging type based not only on the product, but using several aforementioned additional factors relating to a specific shipment of an item to a customer. Additionally, we would like to estimate the causal impact [16] of receiving damage products on customer's spend patterns and factor it into our optimization algorithm.

References

1. Bachu, Y.: Packaging Demand Forecasting in Logistics using Deep Neural Networks. Master's thesis, Blekinge Institute of Technology (2019)
2. Barlow, R., Bartholomew, D.J., Bremner, J., Brunk, H.: Statistical inference under order restrictions; the theory and application of isotonic regression. Wiley, New York, NY, USA (1972)
3. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* **20**(3), 273–297 (1995)
4. Corliss, G.: "Which root does the bisection algorithm find?". *SIAM Review* **19**(2), 325–327 (1977)
5. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* **17**(83), 1–5 (2016)
6. Ivanov, V.: The theory of approximate methods and their application to the numerical solution of singular integral equations. Springer, New York, NY, USA (1976)
7. King, G., Zeng, L.: Logistic Regression in Rare Events Data. *Political Analysis* **9**, 137–163 (2001)
8. Knoll, D., Neumeier, D., Prüglmeier, M., Reinhart, G.: An automated packaging planning approach using machine learning. *52nd CIRP Conference on Manufacturing Systems* pp. 576–581 (2019)
9. Li, L., Lin, H.T.: Ordinal Regression by Extended Binary Classification. In: *19th International Conference on Neural Information Processing Systems*. pp. 865–872. MIT Press (2006)
10. M.-Machuca, M., Sainz, M., M.-Costa, C.: A Review of Forecasting Models for New Products. *Intangible Capital* **10**(1), 1–25 (2014)
11. Mao, F., Blanco, E., Fu, M., Jain, R., Gupta, A., Mancel, S., Yuan, R., Guo, S., Kumar, S., Tian, Y.: Small boxes big data: A deep learning approach to optimize variable sized bin packing. *CoRR* **abs/1702.04415** (2017)
12. Martínez, M.T.A.: Models and algorithms for solving packing problems in logistics. Ph.D. thesis, Universidad de Castilla - La Mancha (2015)
13. Oneto, L., Ridella, S., Anguita, D.: Tikhonov, Ivanov and Morozov Regularization for Support Vector Machine Learning. *Mach. Learn.* **103**(1), 103–136 (2016)
14. Papadimitriou, C.H., Steiglitz, K.: *Combinatorial Optimization: Algorithms and Complexity*. Dover, Mineola, NY, USA (1998)
15. Platt, J.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers* **10**(3), 61–74 (1999)
16. Rubin, D.: Causal Inference Using Potential Outcomes. *J. Amer. Statist. Assoc.* **100**(469), 322–331 (2005)
17. Tikhonov, A., Arsenin, V., John, F.: *Solutions of ill-posed problems*. Winston, Washington, DC, USA (1977)
18. Wilson, R.C.: A packaging problem. *Manage. Sci.* **12**(4), B135–B145 (Dec 1965)