塑料 Global Plastic Waste Model based on Linear Programming and Support Vector Regression

摘要

Plastic waste (PW) has been one of intractable environmental dilemma for modern humankind. In order to mitigate this tough problem, we combine production of plastics and management of plastic waste to establish a model, which can evaluate the impacts of plastic waste and find out an achievable minimal level of plastic production that can be reduced with a time line.

Firstly, we develop a plastic waste estimate model based on Product Life Cycle (PLC) theory to calculate the maximal levels of single-use or disposable plastic product waste within the environmental capacity in a certain region or country. Then we independently explore the four possible ends of PW and formulate their impacts on environment by several functions. We use linear programming that includes one particular objective function and four constraints. As a result, we can output the maximal amount of plastic products by analyzing the input digits of any region.

Secondly, to find out to what extent plastic waste can be minimized, we established an model called HSVR which combines happiness-index analysis and SVR method. By analyzing the characteristics of different regions and people＇s living standards, we find out the relationship between the amount of plastics produced and people＇s happiness index. Based on this, we can analyze how much plastic production we can reduce without significantly reducing happiness.

Thirdly, we follow the previous happiness index to assess how much human life is changed, and apply the Environmental Assessment of Solid Waste Systems and Technologies (EASEWASTE) model to evaluating how the the environment is affected by plastic reduction. Moreover, we take computable general equilibrium (CGE) model to quantify the effects of the target plastic production. Then we normalized the outputs and into non-dimensional indices, and calculate the weight for each indicator above by the analytic hierarchy process (AHP). The total impact for achieving the confined level is a linear combination of the three indices.

Furthermore, reduction of plastic production will impact countries unequally because of their different development level. We use dual programming to fine the dual model of the linear programming based model we established first. Then shadow prices of every factor formulated by constraints can be calculated, which indicate the significance of each factors. Then we divided them into environmental and governmental factors. Sensitivity analysis is be used to identify which factors that is most sensitive to a particular country, and the equitable reduction of plastic of the country depends on the categories that its sensitive factors belong to.

Finally, we analyze how to approach the minimum level following the time line by an achievable way. Many unexpected circumstance would definitely appear so we empirically choose most

significant possibilities that may delay or accelerate the achievement process. The ultimate result are specified into a memo which will be provided for ICM.

目录

1 Introduction

1.1 Background

对于亚马逊等电商平台来说，最不缺的就是顾客的评价和评论，评级通常用 1 到 5 星五个等级来表达，而文本评论让客户有了表达更多想法和反馈的机会。但是如何利用好这庞大的评价和评级数据却也是个不小的挑战。运用好这些数据不仅能帮助商家针对性地提供客户服务，还能发现潜在的高销量以及低销量商品，进行运营策略的调整，或者及时地改善产品。

1.2 Problem statement

现在，阳光公司打算在线推出三款新的产品，分别是：微波炉，婴儿奶嘴以及吹风机，并获取了客户的历史评级和评论。他们希望从评论，评级中挖掘出有效的信息，帮助公司更好地制定销售策略，增强产品的效用。他们希望以下的问题能得到合理的解决：

1. 产品的星级，评论和帮助评分之间到底有什么样的关系? 能否把评论量化，探究三者之间的数学联系?

2. 怎样建立一个模型能综合评级和评论两个因素，并帮助公司获取有效信息，调整营销策略?

3. 是否能从时间的角度描述产品的信誉，从而能看出产品在在线市场上的声誉变化?

4. 根据文本评价和评级，我们怎么才能发现潜在的成功或失败的产品?

5. 特定的星级与评论数之间是否有关联? 换句话说什么样的评级更能吸引客户进行评论?

6. 文本评价的主题词例如：'高兴'，'失望' 是否和特定的评级紧密相关?

2 Model Analysis

2.1 Analysis of task 1

2.2 Analysis of task 2

2.2.1 Rating-review based reputation model

我们可以从产品的星级和评论推断出产品的声誉，而一个产品的声誉越好往往意味着产品的潜在销售量越高，在这里，我们不能单纯地用星级或评价衡量一个产品的声誉，因为生活中往往存在着"五星好评"和"一星黑粉"的情况。因此我们综合运用星级和评论构造出声誉的模型，通过比对产品的声誉，阳光公司可以针对性提升声誉高的产品的备货量和宣传力度。

同样，声誉的计算需要先对某一产品的评论进行 LDA 处理，转化为离散的等级值, 为了考虑评论的可信度，我们还考虑了 vine, verified_purchase, helpful_votes, tot_votes 几个字段，这些字段可以对评论的准确度进行衡量，因为购买过产品的评论才更可信，helpful votes 高的评论才更具有参考性。

通过建模，并对 2015 年 7 月-8 月评论数大于 25 的 pacifer 样本进行计算，我们发现，在 7-8 月，样本'wubbanub soft toy and pacifier' 的声誉更高，销售前景最大；样本'philips aventbpa free soothie pacifier, 3+ months, green, 2 count' 的声誉最低，销售前景最差。

2.2.2 Relations between time and reputation

产品的信誉会随着时间的变化而波动，而其中潜在的规律可能会对阳光公司有很大的启发。我们可以通过 RBF 神经网络 [22] 对 (信誉-时间) 作函数拟合，在得到拟合曲线后，就可以直观地观测到信誉的变化趋势。以奶嘴为例，

- 首先，我们挑选奶嘴中评论数最多的样本–'philips avent bpa free soothie pacifier, 0-3 months, 2 pack'(样本数为 833)

- 然后利用信誉转换模型得到所有样本的信誉

- 接着以一天为单位时间，利用 RBF 神经网络，对信誉与时间作拟合

- 最后，根据拟合曲线，我们发现 philips 奶嘴的信誉在 2015 年 5 月 30 日有极小值，在 5 月 10 日有极大值，并在下一问中对出现极大、极小值的原因进行了分析。

2.2.3 Indication of potentially successful or failing product

正如上文中提到的信誉会随时间波动，那么我们拟合的曲线自然会有拐点。因此我们定义如果曲线将迎来极大值，那么这个产品将成为潜在的成功产品；如果曲线将迎来极小值，那么这个产品将成为潜在的失败产品。

为了方便处理，我们建模时只考虑曲线的一组极大和极小值，可以通过对曲线极大、极小值附近的星级、评论进行聚类，找到词频最高的主题词与星级。当这些主题词和星级出现的时候就意味着很可能产品的声誉将迎来拐点。

在我们的实验中发现当 xxx 和 xx 出现的时候，产品的声誉将下降，那么这个产品会成为潜在的失败产品; 当 xx 和 xx 出现的时候，产品的声誉将上升，这个产品会成为潜在的成功产品。

2.2.4 Relationship between specific star ratings and more reviews

Whether a particular star rating will cause more comments. We can judge this is a statistical problem and one-way ANOVA can be used. We need to count the average star rating in the previous period and the total number of comments in the next period to find their relationship. Take pacifier for example.

- First of all, we selected the most commented samples in pacifier, which is 'philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary'(833 samples).

- Then we count the number of comments between different time periods of the product and the average star rating before the time period.

- Because the independent variable only has star rating, we discretize it into high star rating, medium star rating and low star rating. Then we use one-way ANOVA to find the relationship between star ratings and reviews.

- Finally, according to the statistics, we found that The relationship between the average star rating in the current time and the total number of comments in the next time period is not strong, but after the semantic analysis of the comments, we divide the comments into positive comments, medium comments and bad comments. We found that in the face of a period of time when products are low stars, users are more willing to give negative comments; in the face of a period of time when products are high stars, users are more willing to give positive comments.

2.2.5 Relationship between specific quality descriptors of text-based reviews and rating levels

The topic asks us to find out whether the specific words appearing in the comment, such as "disappointed", "enthusiasm", will be related to the rating of the comment. Because we are looking for the relationship between the data, we can determine whether this is a statistical problem. We can still use one-way ANOVA, taking the blower as an example.

- First of all, we make word frequency statistics for the comments in the hair dryer, and select the words with higher frequency.

- Then, we select the comment with key words in the hair dryer. We code the comment 0 or 1 based on whether it contains key words or not.

- Then because the independent variable only contains keywords, we continue to use one-way ANOVA to find the relationship between keywords and star ratings.

- Finally, according to the statistics, we found that comments with the keyword "disappointed" are closely related to star rating of one star, and comments with the keyword "love" or "great" or "good" are closely related to star rating of five stars.

3 Assumptions and justifications

To simplify the problem, we strive for the following assumptions:

- Users can see the rating stars of previous users before commenting, but can't see the rating stars of users long ago. That is, only the average star rating of the previous time period can affect the comments of the next time period..

- Comments on emotional tendencies are consistent with star ratings.

- Star ratings in a time period satisfies the normal distribution.

4 Parameter Table

Important symbols that in this article are listed in Table 1.

表 1: Notations

Symbols	Definition
s_i	Annual certain plastic production.
N_i	The number of $CO2$ generated by the combustion of 1 molecule specific plastic unit.
EA	Emission to air.
EW	Emission to water.
EC	Emission of CO_2.
TEA	Total emission to air.
TEW	Total emission to water.
TEC	Total emission of CO_2.
CEA	Compensatory emission to air of recovery process.
CEW	Compensatory emission to water of recovery process.
SM	Plastic waste into marine.
M	The quantity of annual management of marine plastic.
V_air	Total available air of a specific country/region.
V_water	Total available water of a specific country/region.
σ	Proportion of incinerated plastic waste.
μ	Compensate.
θ	The percentage of GDP plastic industry contributes, quantified by coefficient.
ι	Tossil fuel consumption percentage. Generally, ι equals 4%.
k	Proportional coefficient of carbon emission per unit of
$\overline{s_j}$	The length of the business cycle
$\overline{s_j}$	The average star in the j-th business cycle
C_{j+1}	The total number of comments in the (j+1)-th business cycle energy consumption

5 Model Construction

5.1 Rating-review based reputation model

通过对评级和评论建模可以得出产品的信誉，而更高的信誉意味着更高的潜在销量，可以帮助阳光公司作出更好的决策。建模过程如下：

假定产品的星级为 R_α，评论量化之后为 R_β，评论的 helpful votes 为 R_h，则 unhelpful votes(R_u) 为 total votes 减去 helpful votes，净投票数为（$R_h - R_u$），我们用净投票数加 1 作为因子再乘上量化后的评论，表征评价的可信度。

我们采用对数形式使得中间计算更平滑，然后对该时段的信誉进行平均，最后参考文章 [25] 用

指数形式对结果差异进行放大, 信誉模型初步建立为:

$$F_{repu} = e^{\frac{\sum_{i=1}^{n}(log_2(R_\alpha+1)+\beta \times log_2(R_\beta \times (R_h - R_u+1)+1))}{n}}$$ (1)

其中 n 为某时段评价数。为了方便处理，在模型中，当净有用投票数为负数时，统一设为-1，即这种评论对声誉没有贡献。

更进一步，超参数 β 一方面要能体现评级和评论的权重关系，另一方面还要能表征评论的可信度，而 vine 或者 verified_purchas 字段为'Y' 的评论是体验过产品的评论，更具有参考性。因此 β 的表达式如下:

$$\beta = \begin{cases} 1 + W_r & if \quad vine \quad or \quad verified\_pruchase =' Y' \\ W_r, & else \end{cases}$$ (2)

其中 W_r 为上文中提到的评级和评论的关系系数。此处为 0.61.

为了检验我们的模型，我们先统计了从 2015 年 7-8 月的所有评论数在 25 条以上 pacifier 样本，分别为:

- philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary, 标注为 Sample1

- wubbanub soft toy and pacifier，标注为 Sample2

- wubbanub brown monkey pacifier, 标注为 Sample3

- wubbanub infant pacifier - giraffe, 标注为 Sample4

- philips aventbpa free soothie pacifier, 3+ months, green, 2 count, 标注为 Sample5

- wubbanub brown puppy pacifier, 标注为 Sample6

然后我们对 6 种样本计算了这两个月内的平均声誉值，结果如图:
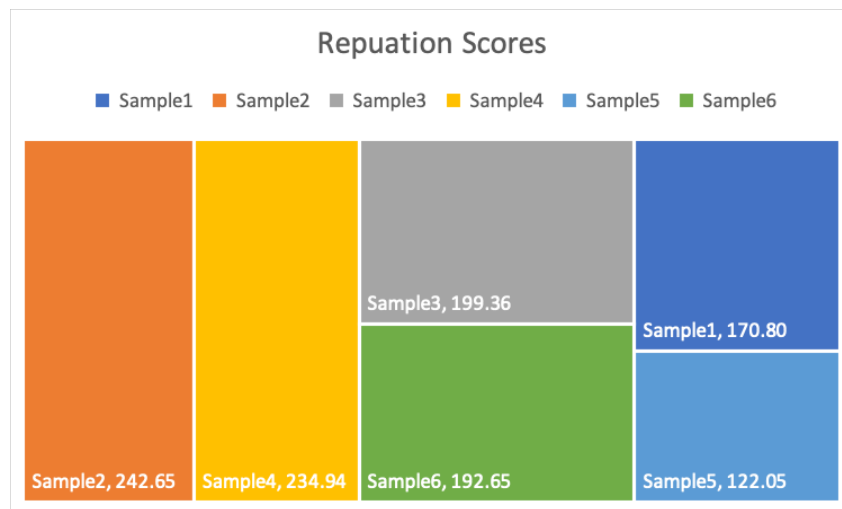
图 1: The reputations of six samples calculated from July to August 2015

图中说明，2015 年 7 月-8 月 pacifier 中的 Sample2、Sample4 属于同一档，产品声誉值较高，意味着更好的销售前景; 其次是 Sample3 和 Smaple6 属于同一档; 最后是 Sample1,Sample5 属于同一档，声誉值较低，意味着销售前景并不明朗。事实上，我们发现 Sample5 的评级里确实充斥着 1 星，2 星，评论中也不乏许多差评。

5.2   Reputation-time model based on RBF regression

探究产品信誉随时间的变化规律有助于阳光公司调整策略，增加收益。受到 [24] 的启发，我们可以采用神经网络进行信誉的拟合及预测。

此处我们用 RBF 神经网络作信誉-时间的回归拟合，因为 RBF 具有很强的非线性拟合能力，可映射任意复杂的非线性关系，而且学习规则简单，便于计算机实现。我们的 RBF-based reputation-time model 建立过程如下：

1. 我们先对数据进行筛选，选择每类产品最具代表性的样本:

   - Pacifier: 'philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary'，有 833 个样本, 分布在 557 个日期。

   - Hair Dryer: 'remington ac2015 t|studio salon collection pearl ceramic hair dryer, deep purple'，有 587 个样本, 分布在 436 个日期

   - Microwave: 'danby 0.7 cu.ft. countertop microwave'，有 394 个样本，分布在 224 个日期

2. 将筛选出的样本代入信誉模型，得到样本信誉随时间变化的离散序列。

3. 搭建 RBF 网络：

   (a) 将样本的前一半作为训练集、后一半作为测试集，并作归一化处理

   (b) 设定网络的神经元数逐步增加, 最多就是训练样本个数，扩展速度 spread 设定为 0.1，开始训练

   (c) 测试 RBF 网络，得到拟合结果图。

我们得到的 2015 年部分拟合结果图如下所示:
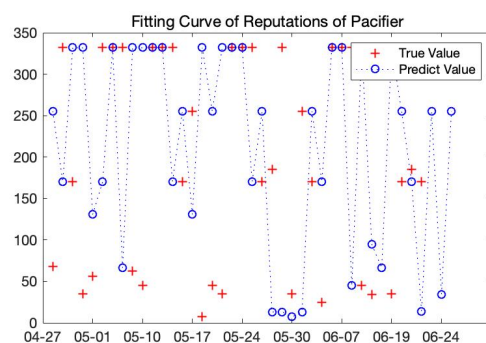
图 2: Fitting Curve of Reputations of Pacifier
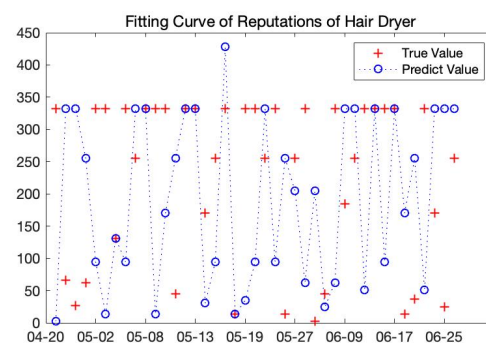
图 3: Fitting Curve of Reputations of Hair Dryer
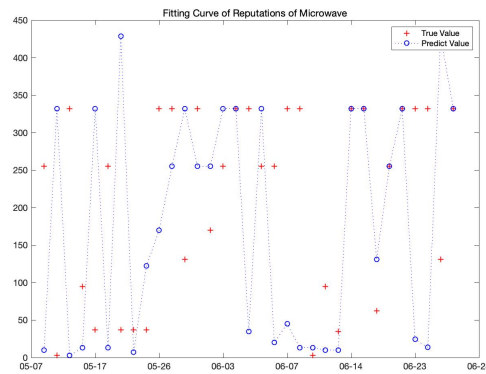
图 4: Fitting Curve of Reputations of Microwave

从图中我们可以观察到以下现象：

- 对于 Pacifier 信誉极小值出现在 2015 年 5 月 30 号左右，Hair Dryer 信誉极小值出现在 2015 年 5 月 19 号左右，Microwave 信誉极小值出现在 6 月 7 号左右。出现极小值的原因我们将在下一问中进行阐述。

- 对于 Pacifier 信誉极大值出现在 2015 年 5 月 10 号左右，Hair Dryer 信誉极大值出现在 6 月 9 号左右，Microwave 信誉极大值出现在 5 月 26 号左右。同样，我们将在下一问讨论出现极大值的原因。

5.3 Model to indicate potentially successful or failing product

在我们的建模中，拟合曲线极大值附近常出现的主题词和星级是出现潜在成功产品的标志，而拟合曲线极小值附近常出现的主题词和星级是出现潜在失败产品的标志。

以 Pacifier 为例，根据图??, 拟合曲线极大值出现的时间点为 2015 年 5 月 10 日左右（标记为 t_1），拟合曲线最小值出现的时间点为 2015 年 5 月 30 日 (标记为 t_2)，分别抽取 t_1，t_2 附近 7 个时间点的星级和评论，我们进行 LDA 文本分析并聚类，结果如下：

我们发现，极大值附近，xxx 和 xx 容易出现，这很可能是因为 xxxx；极小值附近，xx 和 xxx 容易出现，而这可能是因为 xxx。

5.4 Model to the relationship between specific star ratings and more reviews

Because there are only specific star ratings can cause more comments from users. Obviously, there is only one independent variable, so we use one-way ANOVA [1]. Our Model is built as follows.

1. We first filter the data and select the most representative samples of each type of product:

   - Pacifier: 'philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary'，833 samples。

- Hair Dryer: 'remington ac2015 t|studio salon collection pearl ceramic hair dryer, deep purple', 587 samples。

- Microwave: 'danby 0.7 cu.ft. countertop microwave', 394 samples。

2. Further, we screen samples to remove comments whose star ratings are inconsistent with comments.

3. Then we split the time into d days (generally 7 days for business) and assume that the comments in the next period are related to the average star rating in the previous period. Next, the average star rating $\overline{s_j}$ of the i product in the j-th time period and the number of comments in the next time period C_{j+1}. We need to mine the relationship between $\overline{s_j}$ and C_{j+1}:

The average star level can be any value between $[1, 5]$. For the convenience of seeking the relationship, we can discretize it by dividing $[1, 5]$ into four sections, each section is corresponding to one class, and then $\overline{s_j}$ is divided into different classes according to its value. So, the question is translated into, whether a specific star rating range corresponds to C_{j+1}. Because we only consider the influence of star rating, we use one-way ANOVA.

One-way analysis of variance requires that all variables follow normal distribution, and the variance is uniform

表 2: Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
2.054	2	812	0.129

From the graph, we find that the variance is homogeneous, and we can use one-way ANOVA：

表 3: ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.31	2	6.15	16.599	0.001
Within Groups	13402.85	812	16.51		

We can see from the picture that:

- The relationship between the average stars in the current time period and the total number of comments in the next time period is not strong

Later, we made a semantic analysis of the comments.

1. According to the results of semantic analysis, we divide comments into positive comments, medium comments and negative comments.

2. Then we calculated the relationship between the proportion of positive comments, medium comments and negative comments and A_p

The final statistical results are as follows:

表 4: Descriptive of the percentage of negative comment

rating	N	Mean	Std.Deviation	Std.Error	F	Sig.
2	2	56.25	8.83	6.25	16.599	0.001
3	18	36.64	42.52	10.05		
4	175	8.68	19.47	1.47		

表 5: Descriptive of the percentage of medium comment

rating	N	Mean	Std.Deviation	Std.Error	F	Sig.
2	2	37.50	17.68	12.50	0.40	0.671
3	18	20.40	27.16	6.40		
4	175	22.32	25.52	1.93		

表 6: Descriptive of the percentage of medium comment

rating	N	Mean	Std.Deviation	Std.Error	F	Sig.
2	2	6.25	8.83	6.25	9.828	0.001
3	18	42.95	40.88	9.64		
4	175	68.99	29.36	2.21		

From the figure we can draw the following conclusions:

- In the face of a period of time when products are low star rating, users are more willing to give negative comments.

- In the face of a period of time when products are high star rating, users are more willing to give positive comments.

5.5 Model to the relationship between specific quality descriptors of text-based reviews and rating levels

To determine whether there is a relationship between keywords and star ratings in the comments, it is obvious that there is only one independent variable, so we still use one-way ANOVA.

Our model to relationship between specific quality descriptors of text-based reviews and rating levels based on one way ANOVA is established as follows

1. We first filter the data and find the specific quality descriptors with high frequency, such as "disappointed", "love", "great".

2. Further, we also filter samples, because some people will add "not" before "disappointed" to make the meaning opposite. So we need to remove these samples

3. Based on whether the comments contain keywords, we divided the samples. And based on this, one-way ANOVA is carried out.

The final statistical results are as follows:

表 7: "disappointed"

	N	Mean	Std.Deviation	Std.Error	F	Sig.
0	11250	4.16	1.27	0.012	745.936	0.000
1	220	1.82	0.84	0.06		

表 8: "love"

	N	Mean	Std.Deviation	Std.Error	F	Sig.
0	9042	3.97	1.36	0.01	553.465	0.000
1	2428	4.65	0.85	0.02		

表 9: "great"

	N	Mean	Std.Deviation	Std.Error	F	Sig.
0	8700	3.99	1.37	0.02	346.118	0.000
1	2770	4.51	0.93	0.02		

These show that the distribution of star ratings are significantly affected by the key words. Now we use histogram to show the specific distribution：
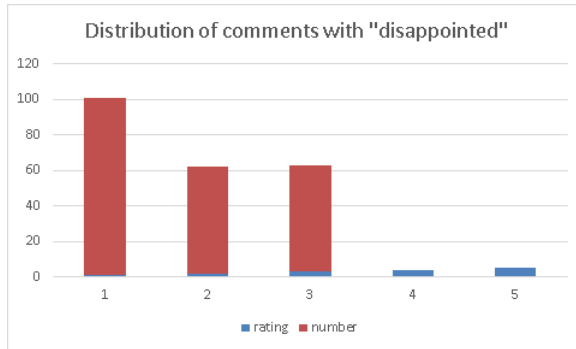
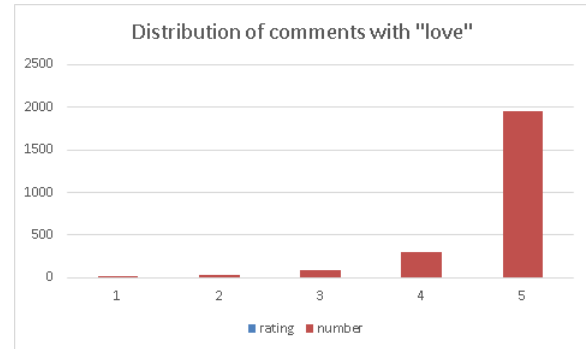图 5: Fitting Curve of Reputations of Pacifier

图 6: Fitting Curve of Reputations of Hair Dryer
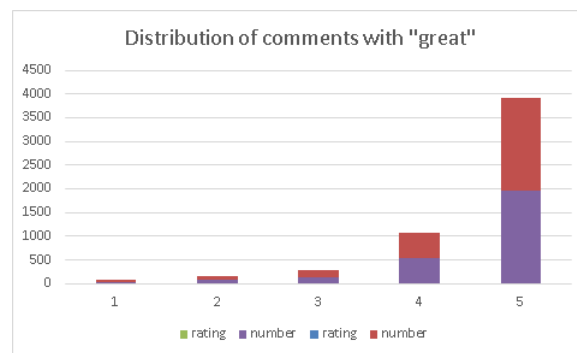
图 7: Fitting Curve of Reputations of Microwave

From the figure we can draw the following conclusions:

- Comments with the keyword "disappointed" are closely related to star rating of one star;

- Comments with the keyword "love" or "great" or "good" are closely related to star rating of five stars.

5.6 Theories related to our model

In this section, we will introduce theories that are closely related to our model.

5.6.1 PLC theory

To guarantee that the maximal usage or producing of plastics will not damage the environment further, we identify the parts of plastics' life cycle that might cause damage to the environment, as is shown in figure 8. Consider a revised product life cycle (PLC) theory: as every product has its life cycle from R&D(esearch and development) to launching on market to decline and exit the market as an end, plastics also cycle from products to waste. For plastic products from various source, their producing processed may affect the environment as well as their managing courses. The model expand based on the plastic life cycle theory.

图 8: The parts of plastics' life cycle that might cause damage to the environment.

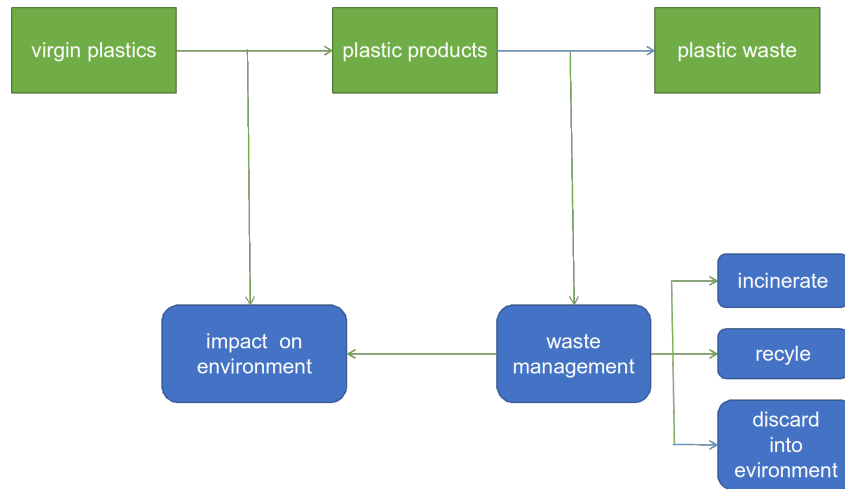### 5.6.2 Linear programming

Linear programming[18] (LP) is a tool for solving optimization problems for which we do following:

1. We attempt to maximize (or minimize) a linear function of the decision variables which is called the objective function.

2. The values of the decision variables must satisfy a set of constraints. Each constraint must be a linear equation or linear inequality.

3. A sign restriction is associated with each variable. For any variable x_i, the sign restriction specifies that x_i must be either nonnegative ($xi \geq 0$) or unrestricted in sign.

So we try to set our objective function as the permitted maximal amount of plastic and abstract the most limited conditions into constraints.

5.7 Qualify the impact on the environment

We analyzed the pollution caused by each stage of the plastic's entire life cycle.

5.7.1 Producing process

According to the first assumption, we consider the annual production and waste as the same variable S. And naturally it is divided into various sorts of plastic.

$$S = \sum_i s_i \tag{3}$$

Note that, i rrepresents a sort of plastic and s_i is the annual production of this sort of plastic. All sorts of plastic taken into account are PVC, PE, PP, PS, PET, PUR and PC[11].

To quantify to what extent plastic waste impact the environment in its life cycle, we use Emissions air(EA) and Emissions water (EW) to measure emissions, whose units are Units of Polluted Water (UPW/ton) and Unit of Polluted Air with the production of a ton of plastic(UPA/ton). UPW is the number of cubic meters of water polluted up to the European drinking water standard by the production of 1 tonne of the material. UPA is a similar measure for air and gives the cubic meters of air needed to dilute the emissions to the European maximum acceptable concentration (MAC)[11]. Then the Emissions of water and air can be formulated below:

$$TEA^\alpha = \sum_i s_i \times EA_i^\alpha \tag{4}$$

$$TEW^\alpha = \sum_i s_i \times EW_i^\alpha \tag{5}$$

Given the energy consumption per unit of plastic production, we set energy consumption of specific plastic production as E_i and the proportional coefficient of carbon emission per unit of energy consumption as k, total emission of CO2 can be calculated as:

$$TEC^\alpha = k \sum_i s_i \times E_i^\alpha \tag{6}$$

5.7.2   Managing process

When it comes to plastic waste management, there are three possible fates for managed plastic waste: being recycled, being incinerated and being discarded(be landfilled or be discarded in natural environment). According to the third and fourth assumption, we assume there are only two managing method: incinerating or recycling.

For the proportion of recycled PW, they are recycled by The material recycling facilities(MRFs). The mechanical separation removes polyethylene terephthalate (PET) and high density polyethylene(HDPE) bottles at high efficiency, which are sent to recycling, and other minor high calorific fluxes sent to energy recovery in cement kilns[4]. After that we could get certain recovery rate, we set this rate as $1 - \sigma$, the proportion of incinerated part is σ.

Firstly, we apply a classical environmental assessment method: life cycle assessment (LCA), to evaluate the environmental issues associated with solid wast with solid waste management[5].

Secondly, we simplify and abstract the many categories of LCA into three categories based on data availability and computational matching: EA, EW and EC.

Furthermore, because the recycling technology(MRFs) saves this part of PW from incineration and put recycled products into market again, while the harmful emissions from recycling are often less than the emissions from incinerating this part of the plastic[15]. There are digital supports that reveal the positive influence about of recovery, as is shown in figure 9 and figure 10. So we calculate this different value as compensatory emissions.

Then the environmental impact of plastic waste management can be calculated:
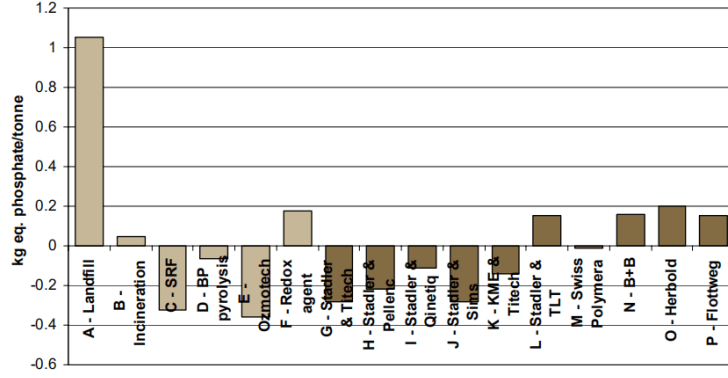
图 9: Net eutrophication potential.
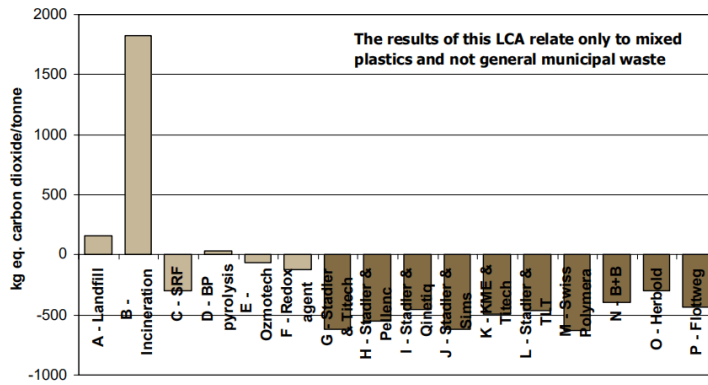
图 10: Net global warming potential.

$$TEA^{\beta} = \sigma \sum_i s_i \times EA_i^{\beta} - (1 - \sigma) \sum_i s_i \times CEA_i^{\beta} \tag{7}$$

$$TEW^{\beta} = \sigma \sum_i s_i \times EW_i^{\beta} - (1 - \sigma) \sum_i s_i \times CEW_i^{\beta} \tag{8}$$

Apart from above analysis, although we assume that all the plastic waste discarded in natural environment will be managed, too much discarded plastic waste (even if temporarily) will notably do harm to marine ecosystem[3]. Thus it's necessary to estimate the mass of plastic waste(symbolized by SM) inputs from land into the ocean, given specific total plastic waste.

$$SM = \varsigma \sum_i s_i \tag{9}$$

For the plastic waste that has been incinerated, in the process of incineration, we take the emission of greenhouse gases CO_2 (EC) to represent it's contributions to global warming:

$$TEC^{\beta} = \sigma \sum_i s_i \times N_i - \mu(1 - \sigma) \sum_i s_i \tag{10}$$

5.8   Constraints given by environmental carrying capacity

Firstly, according to the fifth assumption, the pollutant emission constraints are roughly estimated by the percentage of GDP plastic industry contributes, quantified by coefficient θ.

$$TEA^{\alpha} + TEA^{\beta} \leq \theta V_{air} \tag{11}$$

$$TEW^{\alpha} + TEW^{\beta} \leq \theta V_{water} \tag{12}$$

Secondly, when it comes to greenhouse gas CO_2, it's more propriate to use its fossil fuel consumption percentage to set the restraint because the fossil fuel combustion is the principal factor of the global warming, rather than natural carbon cycles[11].

$$TEC^{\alpha} + TEC^{\beta} \leq \iota V_{CO_2} \tag{13}$$

Last but not least, considering current quantity of plastic discarded in the ocean don't exert severe impact on marine ecosystem, we regard zero net input into the sea an acceptable circumstance for marine environment, i.e. annual plastic waste input can't beyond the quantity of annual management of marine plastic(symbolized by M).

$$SM \leq M \tag{14}$$

5.9   Model paradigm

In the end, our model can be written as follows:

Objective functions:

$$S = \sum_i s_i \tag{15}$$

subject to:

$$\begin{cases} TEA^{\alpha} + TEA^{\beta} \leq \theta V_{air} \\ TEW^{\alpha} + TEW^{\beta} \leq \theta V_{water} \\ TEC^{\alpha} + TEC^{\beta} \leq \iota V_{CO_2} \\ SM \leq M \end{cases} \tag{16}$$

5.10   Analysis of China using our model

In this section, we will combine the data from China to further introduce the model.

China is a developing country which emission largest amount of PW to marine, which may be sensitive to change of plastic amount. As shown in figure 11, China has the most and most

developed plastic industry in the world. We apply our model to the practical examples of China. Table 10 lists seven disposable plastics mainly used in China. We consider them as model variables s_i.

图 11: China's developed plastic industry.

We used data from the National Bureau of Statistics of China[16] and obtained parameters closely related to the model. Specific information can be consulted in the appendix.

Note that for practical reasons, we set some settings for the model, where the compensite emission to water and air of recycling process should be deemed as zero for the following considerations:

1. The proportion of recycled PW is relatively small in realistic practices.

2. Most regions of China are not developed enough to have the advanced technology of MRFs to manage PW.

表 10: The main types of disposable plastics used in China

Variable	Plastic types
s1	PVC
s2	PE
s3	PP
s4	PS
s5	PET
s6	PUR
s7	PC

3. Chinese PW sorting and managing system are not perfect.

Finally, the linear programming model based on Chinese data is as Eq.17 and Eq.18:

$$S = s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 \tag{17}$$

$$\begin{cases} 700s_1 + 265s_2 + 325s_3 + 255s_4 + 180s_5 + 0s_6 + 180s_7 \leq 2.88 \times 10^{11} \\ 3000s_1 + 1650s_2 + 3685s_3 + 6355s_4 + 8000s_5 + 0s_6 + 5050s_7 \leq 7.89696 \times 10^{10} \\ 1.99s_1 + 4.42s_2 + 4.43s_3 + 4.77s_4 + 3.23s_5 + 0.49s_6 + 1.96s_7) \leq 8.062 \times 10^{11} \\ 0.41 \times (s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7) \leq 4.85 \times 10^5 \end{cases} \tag{18}$$

We solved this problem by using the simplex method through MATLAB, and obtained the largest amount of plastic that China can bear in one year. The results show that China can bear 1.18 million tons of plastic a year. When the optimal solution is obtained, the production volume of various plastics is shown in table11.

We will continue to analyze this case later to explore the impact of different constraints.

6   Plastic waste reduced

At present, the global use of plastics has exceeded the upper limit of environmental tolerance, which poses a huge threat to the ecological environment. Considering the huge differences in the existence of countries around the world, we must consider each region separately. In section 6, we propose a model of Happiness based on SVR[6], which we call HSVR. Using HSVR, we will explore the impact of the use of plastic on human living standards in a region, and try to reduce the amount of plastic used as much as possible without affecting much of human living standards.

Our method can be summarized as follows: the use of plastic is essential, and reducing the use of plastic will inevitably affect the living standards of local people. By analyzing the characteristics of different regions (especially the amount of plastics consumed) and the people's living standards, we find out the relationship between the amount of plastics consumed and the people's happiness index.

表 11: The amount of various plastics in the optimal solution

Plastic types	Amount of plastic(t)
PVC	210428.703365915
PE	107339.181740866
PP	174475.173021043
PS	150364.822097472
PET	233445.923823156
PUR	42560.4917899467
PC	264312.533429888

Based on this, we can analyze how much plastic consumption we can reduce without significantly reducing the happiness index.

The following of this section will first briefly introduce the method we use, and then obtain a specific regression model based on the Global Happiness Report[7] and plastic consumption in each region[8].

6.1   SVR algorithm

Support vector regression(SVR) is an application of SVM (support vector machine) to regression problems. Support vector machines construct a hyperplane or a series of hyperplanes in a high-dimensional or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, using a hyperplane to achieve a good segmentation can make the closest training data points have the largest separation distance in any category. This is because usually a larger margin can have a lower generalization error of classifier[9].

Given training vectors $x_i \in R^p, i = 1, \cdots, n$, and a vector $y \in R^n$. $\varepsilon - SVR$ solves the following primal problem:

$$\min_{\omega,b,\zeta,\zeta^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \tag{19}$$

subject to

$$y_i - \omega^T\phi(x_i) - b \leq \varepsilon + \zeta_i \tag{20}$$

$$\omega^T\phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^* \tag{21}$$

where $\zeta_i, \zeta_i^* \geq 0, i = 1, \cdots, n$.

Its dual is

$$\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha - \alpha^*) + \varepsilon e^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*) \tag{22}$$

subject to

$$e^T(\alpha - \alpha^*) = 0 \tag{23}$$

where $0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \cdots, n$, e is the vector of all ones, $C > 0$ is the upper bound, Q is an n by n positive semidefinite matrix, $Q_{ij} \equiv K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$ is the kernel. Here training vectors are implicitly mapped into a higher (maybe infinite) dimensional space by the function ϕ. ϕ can be linear, polynomial, sigmoid or others.

The decision function is:

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + \rho \tag{24}$$

For more details, please refer to [6].

6.2 Detailed analysis

In this section, we will introduce the quantitative standard of happiness, analyze this problem, and use SVR to get the function of happiness and plastic consumption $\hat{f}$ so that $\hat{f} : X \mapsto y$, where X is a vector in $\mathcal{X} = \mathcal{R}^d$, including a dimension of plastic consumption.

6.2.1 The quantitative standard of happiness

We use the happiness quantification standard used by the World Happiness Report[7], where the happiness scores and ratings use data from the Gallup World Poll. The scores are based on answers to the main life evaluation question asked in the poll. This question, known as the Cantril ladder, asks respondents to think of a ladder with the best possible life for them being a 10 and the worst possible life being a 0 and to rate their own current lives on that scale. The scores are from nationally representative samples for the years 2013-2016 and use the Gallup weights to make the estimates representative.

Obviously, people's happiness depends not only on the amount of plastic consumed. In fact, plastic consumption and waste account for only a small part of the factors affecting people's happiness. There are six factors usually concerned to be relevant to happiness: economic production, social support, life expectancy, freedom, absence of corruption, and generosity. The data that estimates the extent to which each of six factors contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world's lowest national averages for each of the six factors, is also used to as part of the input of HSVR, with some preprocessing:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{25}$$

where x_max, x_min is the maximum and minimum of the original data, respectively.

Note that the first preprocessing maps the data to $[0, 1]$. The preprocessing step will erase the differences between the data formats and make the characteristics of the data more obvious.

6.2.2 Plastic Emphasized

The impact of the consumption and use of plastic on happiness should be far less than that of the other six factors mentioned above, which means that the input to the problem should be a weighting of all seven factors:

$$X = (\epsilon_1 x_1, \epsilon_2 x_2, \epsilon_3 x_3, \epsilon_4 x_4, \epsilon_5 x_5, \epsilon_6 x_6, \epsilon_7 x_7) \tag{26}$$

However, what we need to find is mainly the relationship between plastic consumption and happiness, so we give plastic consumption a sufficiently high weight, where $\epsilon_i = 1, i = 1, \cdots, n$.

The lack of data[10] in research on plastic production and consumption has always been a serious problem. To the best of our knowledge, there are no specific statistics on the amount of plastic used in many regions, especially in developing countries. This means that our research will face problems of insufficient data volume and potential data imbalances. Based on this, we chose SVR as our classifier, because SVR has the following characteristics[9] and is suitable for solving this problem:

- Effective in high dimensional spaces.

- Versatile: different Kernel functions can be specified for the decision function.

6.3   Model parameter selection

We randomly selected 20 countries from [7] and [8] as our training set as our training set. These countries include countries that consume more plastic, such as China and India, and developed countries, such as the United States and the United Kingdom. In combination with the training situation, we adjusted the model in real time to obtain the optimal model.

6.3.1   Model performance evaluation

We choose the Euclidean norm of the regression value and the true value(or the mean square error, MSE) to measure the quality of the regression result:

$$MSE = \frac{1}{k} \|y - \hat{y}\|_2 \tag{27}$$

where y is the true value, and $\hat{y}$ is the regression value.

6.4   Specific model parameters

We have implemented four kinds of kernel function[9] of SVR and verified them separately.

The four cores are:

- linear: $\langle x, x' \rangle$

- polynomial: $(y\langle x, x' \rangle + r)^d$.

- rbf: $e^{-\gamma \|x - x'\|^2}$.

- sigmoid: $\tanh(\langle x, x' \rangle + r)$.

Due to the high feature dimension of our training set and the small number of training samples, in general, the effect of the linear kernel should be better. The results in figure 12 also illustrate this:
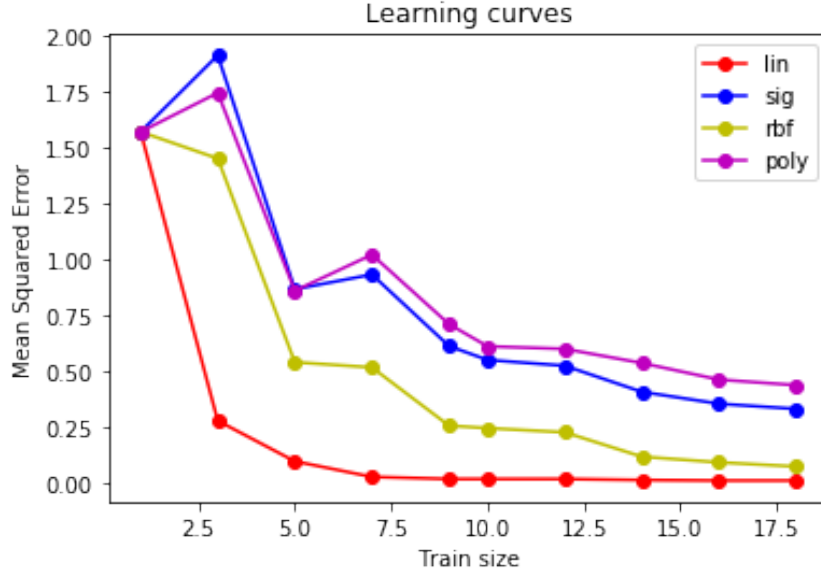
图 12: Training results of different SVR kernel functions.

As can be seen from figure12, the results of linear and rbf are much better than the results of sigmoid and polynomial, which all converge to an error close to 0. In fact, the training error of the linear kernel is on the order of 10^{-2}. Moreover, the linear kernel has already converged by training about ten samples, which shows that our data volume is sufficient for our model.

6.5   Predicting minimum plastic usage

From the above analysis, we find that the linear kernel function has the best fitting effect, indicating that the plastic consumption has a linear relationship with the living standard. This is consistent with per capita plastic consumption in different regions. Table 12 is a table of plastic consumption in different regions. Intuitively, the consumption of plastic is positively related to the living standard of the region. Developed regions generally have higher plastic consumption. Of course, this is also related to other factors. These are reflected in our model.

Based on this conclusion, we found that reducing the amount of plastic used will inevitably affect people's happiness. Therefore, we need to make a compromise between happiness and environment. This is undoubtedly a very painful thing. Therefore, we need to find a threshold on the tolerance of the regions κ. κ is such a value: in a specific area, if happiness $\Gamma = \hat{f}(X) < \kappa$, it is not worthwhile to reduce consumption. Note that $\hat{f}$ is trained in section 6.3. κ can be derived from historical data and is closely related to each place.

Unfortunately, many people are unwilling to pay a lot for environmental protection[14], as is shown in figure 13. So we carefully set this value to 0.5% of the existing happiness index.

For example, figure 14 shows this relationship in the US. Currently, plastic consumption in the United States is 15.719 kg per people. If the amount of plastic used per person per year in the United States is reduced to 6.19 kg per year, the happiness level in the United States will be reduced by 0.5%. We think of this as the largest reduction in plastic use in the United States.

表 12: Plastic consumption in different regions

Area	Plastic consumption (1 ton per 100,000 people)
China	1316.1
North America	1571.9
Asia Pacific	657.1
Western Europe	2870.6
India	251.9
Middle East	580.4
Central and South America	308.8
Central and Eastern Europe	1468.4
Africa	99.8
Japan	812.9
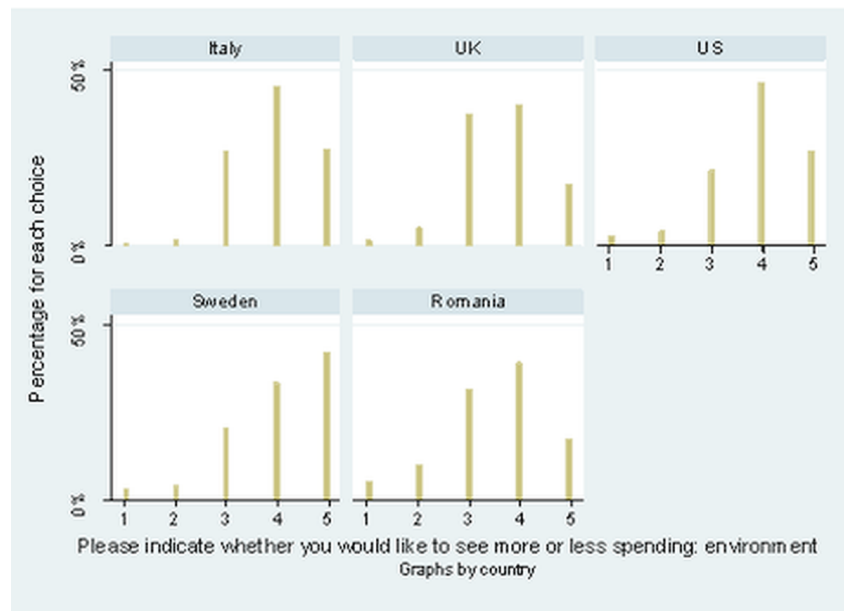
图 13: Happiness-Plastic curve of US.

7 The minimal target of PW and its impacts

Given that we have get the defined maximal amount of palstic production and the most extent that the plastic waste can be reduced to, taking a statistics of all the regions in the world, we can then get the lowest global target. Next, we analyzed the impact of this measure on human life, environment and plastic industry.

7.1 Basic Theories

We have used some theories in this section as a guide, which will be briefly introduced below.

图 14: Happiness-Plastic curve of US.

7.1.1 EASEWASTE

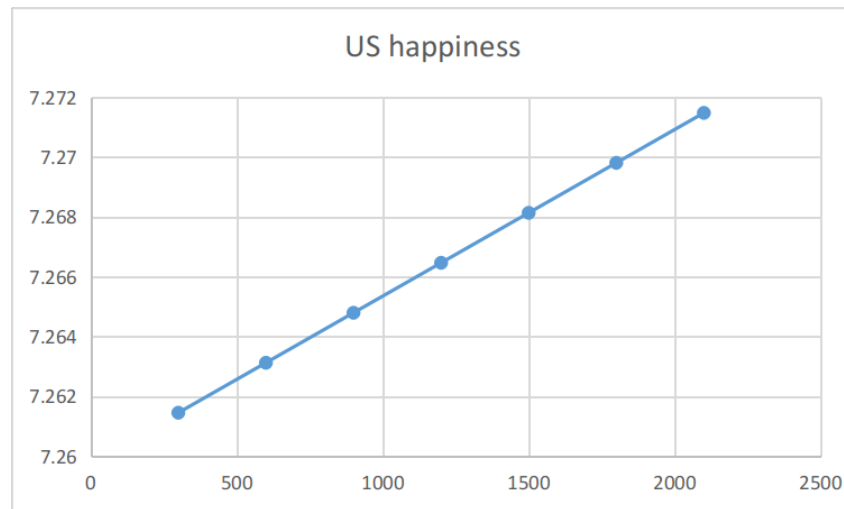TheEnvironmental assessment of solid waste systems and technologies (EASEWASTE) is a model that enables waste managers to assess environmental impacts of solid waste management systems and to make comparisons among recovery technologies for a region with a given population and waste generation. Users can define all necessary data for waste composition, collection, treatment, recovery and disposal and life cycle inventory data for materials and energy used in the waste management system[5].

7.1.2 CGE

The CGE model explicitly assumes that the behavior of all economic subjects is optimized in accordance with the usual neoclassical microeconomic theory, so it is a model concerned with the behavior of general rather than local economic subjects. It uses the assumption that markets are balanced, not unbalanced, and that all markets are settled at the same time. The basic economic units it analyzes are producers, consumers, governments and foreign economies[20].

7.1.3 AHP

AHP has particular application in group decision making and is used around the world in a wide variety of decision situations, in fields such as government, business, industry, healthcare, shipbuilding and education.

Rather than prescribing a "correct" decision, the AHP helps decision makers find one that best suits their goal and their understanding of the problem. It provides a comprehensive and rational framework for structuring a decision problem, for representing and quantifying its elements, for relating those elements to overall goals, and for evaluating alternative solutions[21].

7.2 Global Target

The global goal should be the sum of the small goals in each region. So we divided the world into 10 typical countries according to their plastic assumption amount, as is shown in table 13.

In a certain category, there are similar plastic assumptions and culture background in a certain category.

For each region, we used the model explained in Section 6 to calculate the target for minimum plastic use. Global goals can be achieved by adding them together. All details are shown in table 13.

Note that the total achievable minimal plastic consumtion obtained is about 6.6 hundred million tons.

7.3 Impact Path Analysis

In this section we discuss three most relevant aspects that can be changed: human life, environment and plastic industry.

7.3.1 How human life will be alert

In section 6, we have discussed how people's happiness index will be affected by the reduction of plastic usage. But happiness only reflect people's satisfaction and quality of life, there are still other possible changes of human life.

For example, living style of most people may be changed. Because the supply of plastic is less than the demand, the price of plastic products is bound to rise. People will be pushed to choose products made of other materials as the succedaneum, which may lead to popularity of new materials in the market.

But in the long run, less pollution and more substantial development will lead a better quality of life to every one.

7.3.2 How environment will be affected

It should be noticed that there is a 2.2 hundred million difference between the maximal and minimal amount of plastics. Since there are part of plastic plastics that should be recycled for

表 13: Estimated minimum usage for every region.

Region	Polulation	Minimum plastic	Estimated minimum usage
China	13.86	425	5890.42
North America	3.70	619	2290.49
Asia Pacific	7.12	108	756.01
Western Europe	1.41	1975	2784.72
India	13.39	2700	37087.65
Middle East	4.90	1875	9188.28
Central and South America	6.30	1050	6614.18
Central and Eastern Europe	1.21	620	744.01
Africa	12.16	49	596.05
Japan	1.28	230	291.62

realizing the target reduction. Based on LCF thoery, we can use EASEWASTE theory to estimate the positive externalities on the environment of the recovery. The environmental impacts are calculated based on the overall environmental exchanges which comprise resource consumptions and emissions to air, water and soil which in the life cycle impact assessment may contribute to environmental impact potentials[5].

In section ??, we have quantified the negative impacts of plastic production or plastic waste under quite a few assumptions. When the plastic waste levels are reduced to target values, the emission harmful gases and carbon dioxide will decline proportionally. The pollution caused by this 2.2 hundred million plastic waste will also be migrated.

7.3.3 How to influence plastic industry

This is one of the few negative effects of this achievement. Some skeptics insist that once limite the production of plastics, the muti- trillion-dollar plastic industry will significantly shrink, which is bad for economic balance especially for countries that developed on primary and secondary industries like China. The accounting loss of a certain country or region could be calculated as:

$$L = O \times P \tag{28}$$

where L refers to loss, O is annual output of plastics, and P is annual industory average price of plastics.

Additionally, we should also consider the economic loss or benefit which enrolled the opportunity cost. Computable General Euilibrium(CGE) provide us a fair method to analysis the general influence of this reduction to the plastic industory or even the whole economic.

CGE suggest that when plastic market come to an equilibrium again after the reduction:

1. Producer will switch to other industries and may lead to prosperous of some new and clean materials, which will improve econommic growth and productivity release.

2. Consumers will be willing to change their consuming habitats and more likely to accept new products, which will boost domestic demand and stimulate consumption.

3. The government can put up new policies to reduce the plastic production by adjusting taxes and import and export quotas, which may adjusts the tax and tarrif policy and make more fiscal revenue.

7.3.4 Total impact

In general, the total impact should be a linear combination of the effects of each factor:

$$I = \sigma_1 i_1 + \sigma_2 i_2 + \sigma_3 i_3 \tag{29}$$

where σ_1, σ_2 and σ_3 are the linear coefficients of different factors.

AHP is used to list the matrix and determine the weight of each factor on the whole society. Since the above three factors are abstract social phenomena, we can first quantify the impact with some indicators. Since quantized results have different orders of magnitude and units, we normalize these effects. Multiply the result by the weight of each factor to get the total impact.

Using AHP, we can determine the value of σ_1, σ_2 and σ_3 so that we can calculate the overall impact of different regions. Generally speaking, the weight of different influencing factors in different regions should be different.

8   Discussion about equity

The plastic problem is a global problem and requires international cooperation to resolve it. However, the situation is different in different countries or regions, and we cannot generalize. Different countries have different industrial structures and are affected to varying degrees by the reduction in plastic use. Intuitively, the more a country is affected by plastic, the more we should tolerate this country and allow them to reduce the use of plastic at a slower rate. Therefore, in this section, we further analyze the model proposed in section ??, and give specific methods to analyze the factors most affected by plastics in a region. Based on this method and continued analysis on Chinese cases, we further give theoretical guidance and relevant suggestions.

8.1   Shadow price

We use shadow prices to evaluate the sensitivity of linear programming models. A shadow price is commonly referred to as a monetary value assigned to currently unknowable or difficult-to-calculate costs. It is based on the willingness to pay principle —in the absence of market prices, the most accurate measure of the value of a good or service is what people are willing to give up in order to get it[17]. Shadow price reflects the value of the optimal use of resources. In the model in Section ??, we measured four resources: water quality, air quality, carbon emissions, and marine debris recycling. The shadow prices of these four resources reflect the impact of the region's reduction in plastics on these four resources.

In our model, resources fall into two categories:

- Natural resources. This type of resource is determined by the region itself, and we cannot change the total amount of this resource in a certain region, such as air and water. If the shadow price of this type of resource is high, then we have no choice but to tolerate this area, because forcing such areas to reduce the production and use of plastic will inevitably cause a great burden on the local area.

- Controllable resources. This is a resource that can be determined by the actions of a country, such as carbon emissions and the amount of marine debris recovered. If the shadow price of this type of resource is high, we must ask the country or region because it is the inaction of this country that has caused the pollution of plastic waste.

For plastic production in a country or region, empirical conclusions often look like this: Developed countries often have better management of plastics and should pay more attention to the

issue of natural resources. Developing countries need a lot of resources for development and they should be required to strengthen their management of plastics.

However, this conclusion is not widely applicable, and there are always exceptions. Therefore, for a specific region, the best way is to calculate the shadow price of each resource, so as to truly determine the impact of each resource.

8.2 Determining shadow prices using dual theory of linear programming

Two dual linear programming problems can be described like this[18]:

$$\max z = CX \tag{30}$$

subject to:

$$\begin{cases} AX \le b \\ X \ge 0 \end{cases} \tag{31}$$

and

$$\max \omega = b^T Y \tag{32}$$

subject to:

$$\begin{cases} A^T Y \ge C^T \\ Y \ge 0 \end{cases} \tag{33}$$

where $C = (c_1, c_2, \cdots, c_n)$, $X = (x_1, x_2, \cdots, x_n)^T$, $y = (y_1, y_2, \cdots, y_m)^T$, A is a matrix with m rows and n columns, $b = (b_1, b_2, \cdots, b_m)^T$.

The shadow price of resources is the dual solution of the linear programming problem Y^* for the following properties:

- Weak duality. For any feasible solution of the original problem X and any feasible solution of the dual problem Y, there are $CX \le b^T Y$.

- Complementary relaxation. If $y_i^* > 0$, then $\sum_{j=1} n a_{ij} x_j^* = b_i$; If $\sum_{j=1} n a_{ij} x_j^* < b_i$, then $y_i^* > 0$.

Based on this conclusion, we can get the following inference: while obtaining the maximum value of the objective function, the Lagrange multiplier[19] corresponding to each constraint is the shadow price of the resource corresponding to this constraint. Therefore, we only need the Lagrange multiplier for the optimal solution to obtain the shadow price.

8.3 Shadow price of China's plastic constraint equation: an example

In this section, we further analyze the constraint equation of plastic consumption in China Eq.17 and Eq.18. Based on this, we give the best way for international cooperation for China.

We use MATLAB to obtain the Lagrange multiplier of Eq.17 and Eq.18 when the optimal solution is obtained, as shown in table 14.

As can be seen from the table 14, the shadow price of China's marine garbage recycling volume is much higher than the shadow prices of other influencing factors. This means that the amount of plastic that China can use is actually largely constrained by the amount of marine debris recovered. Part of China's plastic has entered the ocean, but not a lot of it has been recycled. This is an urgent problem. In fact, our model shows that once the constraint is removed, the amount of plastic that can be used normally in China will be much larger than the existing more than one hundred piles, reaching tens of millions of tons. Such a reduction is actually an acceptable burden for China as is analysed in section 6.

In terms of natural resources, air has the largest shadow price. For China, priority should be given to reducing the production and use of plastics with more air pollution as much as possible, which will have the most obvious ecological benefits to the region.

9 Sensitivity Analysis

9.1 Hyper-parameter β

在我们的声誉模型中，超参数 β 被设定为与关系系数 W_r 相关，并且我们认为评论比评级更重要 [26]，尤其是体验过产品后写下的评论，因为这是人们经过仔细思考后写下的感受，更能反映客户体验。因此，在我们的模型中 β 会增大评论的权重。

事实上，我们发现我们的模型对 β 具有一定的鲁棒性，即 β 的轻微波动不会影响声誉的最终结果：

我们分别把 β 公式 Eq.2中的 W_r 从 0.61 换成了 0.2 和 1.2，重新对 section 5.1中的样本进行了计算，结果如下：

表 14: Shadow prices of Chinese resources

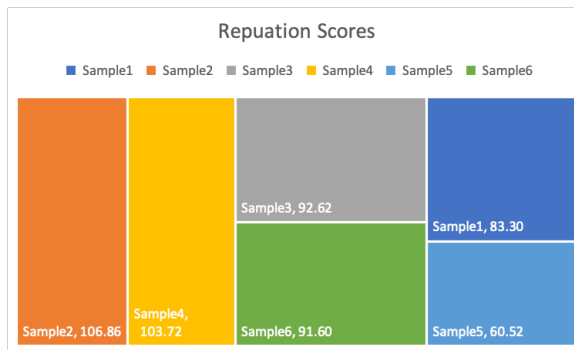| Limiting factor	Shadow price
Air	$1.63233961552303 \times 10^{-14}$
Water	$1.12005836576517 \times 10^{-17}$
Carbon emission	$5.66607374733818 \times 10^{-16}$
Marine garbage recycling	2.43902439024392

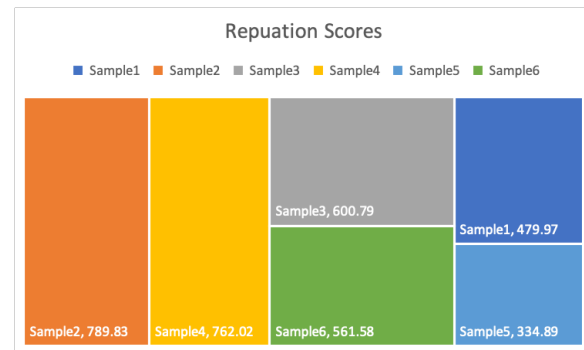图 15: The reputations of six samples calculated from July to August 2015 with lower β

图 16: The reputations of six samples calculated from July to August 2015 with higher β

对比 Fig.1, Fig.15和 Fig.16我们可以看到虽然声誉值的绝对大小改变了，但相对大小没有改变。Sample2, Sample4 仍旧是第一档声誉的产品，Sample5 依旧是样本中声誉处于最差一类的产品。

10   Strength and weakness

The strengths and Weaknesses of our model are summarized as follows:

10.1   Strengths

Model rationality. 我们的模型选择了样本量最大的样本进行分析，可以得出最具有统计意义的结论。对于含虚拟变量的相关性，我们使用单因素方差分析模型，采用了 7 天作为一个商业周期，合理地划分了时间段。同时我们的模型不仅挖掘了星级对于评论的数量影响，还得到了低星级和高星级与评论语义的关系，且该关系较为合理。

Agreement between experimental data and real data. In the process of estimating maximal plastic waste, our model based on PLC theory carefully consider the environmental impact of plastic in different phases of its life cycle rather than merely in the phase of management. What's more, despite the lack of data, our machine learning model still fits the data well. The results of tests on real data sets also validate the performance of our model, where experimental data and real data agrees well.

The amount of data. In the process of modeling, we extensively collected relevant industry data, making the model sufficiently universal and extensible. To our knowledge, few studies on the relationship between plastics and the environment have been able to model using so much data.

10.2   Weaknesses

Data amount of the model. 样本量不太够，导致对于我们对星级的划分不够细腻，导致中等星级与评论间的关系不明显。

Macro analysis. Our model is too detail-oriented and has insufficient performance in macro analysis. In global research, our model encountered some difficulties and showed some uncertainty. In the future, we will try to explore the relationship between the overall situation and the details, and strive to let our model to show the superiority in global problems, too.

11　Conclusion

To mitigate the plastic waste problem, we establish a linear programming base model to estimate the environmental impacts of plastic waste, and get the maximal plastic waste amount without further damage to environment. It is proved that the most extent plastic can be reduced to could be confirmed by HSVR model. A minimal target amount of plastic usage is achieved by expending HSVR model. The equity issue can also be solved by dual programming and sensitivity analysis of factors. Finally, we reckon a timeline to approach this realistic minimal target of plastic using amount.

Letter to Sunshine Company

To: Marketing Director of Sunshine Company

From: Team 2013573

Subject: A Wealth of Data

Date: 2020 年 3 月 9 日

Dear Director:

非常感谢您的信任，雇佣我们团队为您解决这一重要问题。众所周知，产品的评级和评价中蕴含着诸多宝贵的信息，我们希望我们团队从 pacifier, microwave, hair dryer 这三类评价和评级数据中挖掘出的信息，能帮助贵公司作出合理的销售策略调整以及产品的及时完善，希望我们提出的建议能对贵公司有所裨益。

第一问的分析与结果结论，以及意见.

对于如何综合产品的评级和评价为贵公司提供有效信息，我们团队先是对文本评论进行了有效地量化，然后充分利用 helpful votes, vine, verified purchase 等数据甄别了评价的可信度，同时用平均、对数平滑以及指数放大差异等思想，建立了一个较为鲁棒的模型，因为我们的模型总能给出各产品之间准确的相对信誉大小。我们相信，一个好的声誉能潜在地为一个产品带来更好的销量，因此我们建议贵公司对 pacifier, microwave, hair dryer 等产品作信誉评估，然后针对性地增加声誉高的产品的数量以及增大相应的宣传。就像我们论文中提到的一样，对于 2015 年 7-8 月的奶嘴产品，wubbanub infant pacifier - giraffe 是上线销售首选。

我们的拟合模型运用了强大的 RBF 神经网络，较为成功地拟合了信誉同时间的关系，这能帮助我们更好地总结和预测产品的信誉变化，以及从极大极小值中挖掘产品成功或失败的潜在因素，正如我们在文中提到的 xxx 词和 xx 评级出现的时候，我们要格外小心，这是产品的声誉会开始下降，需要贵公司及时完善产品。而当 xxx 词 xxx 评级出现的时候，贵公司可以放心地增加该产品的库存，因为该产品即将热销。

Our statistical model uses one-way ANOVA. After analyzing the average star rating of products in the past period of time and customers' comments of products after that, we find that when customers see products with lower star rating, they are more willing to publish negative comments; When they see products with higher star rating, they are more willing to publish positive comments. Therefore, we suggest that your company should try its best to improve the products before launching the products, so as to avoid more bad comments caused by low star rating, which will lead to a vicious circle.

Our statistical model also finds that some words in the comments are related to specific star ratings. For example, the comments with "disappointed" are closely related to the star rating of one star. For example the comments with "love" and "great" are closely related to the star rating of five

stars. Therefore, we suggest that companies should focus on extracting comments with emotional words, research and analysis, and optimize products in a timely manner.

如果我们的建议能对贵公司有所裨益，我们会非常欣慰。Wish the given advice above would contribute to your work!

参考文献

[1] A. Ross and V. L. Willson, "One-way anova," in Basic and advanced statistical tests. Brill Sense, 2017, pp. 21–24.

[2] Giacovelli, Claudia. "Single-Use Plastics: A Roadmap for Sustainability." (2018).

[3] LI, Wai Chin, H. F. Tse, and Lincoln Fok. "Plastic waste in the marine environment: A review of sources, occurrence and effects." Science of the Total Environment 566 (2016): 333-349.

[4] Rigamonti, Lucia, et al. "Environmental evaluation of plastic waste management scenarios." Resources, Conservation and Recycling 85 (2014): 42-53.

[5] Kirkeby, Janus T., et al. "Environmental assessment of solid waste systems and technologies: EASEWASTE." Waste Management & Research 24.1 (2006): 3-15.

[6] Smola, Alex J., and Bernhard Schölkopf. "A tutorial on support vector regression." Statistics and computing 14.3 (2004): 199-222.

[7] World Happiness Report. https://kaggle.com/unsdsn/world-happiness. Accessed 15 Feb. 2020.

[8] "Plastic Industry Worldwide." Statista, https://0-www-statista-com.lib.rivier.edu/study/51465/global-plastics-industry/. Accessed 15 Feb. 2020.

[9] 1.4. Support Vector Machines —Scikit-Learn 0.22.1 Documentation. https://scikit-learn.org/stable/modules/svm.html. Accessed 15 Feb. 2020.

[10] Groot, Jim, et al. "A comprehensive waste collection cost model applied to post-consumer plastic packaging waste." Resources, Conservation and Recycling 85 (2014): 79-87.

[11] Andrady, Anthony L., ed. Plastics and the Environment. John Wiley & Sons, 2003.

[12] "Cross-Validation (Statistics)." Wikipedia, 7 Feb. 2020. Wikipedia, https://en.wikipedia.org/w/index.php?title=Cross-validation_(statistics)&oldid=939673086.

[13] McLachlan, Geoffrey J., Kim-Anh Do, and Christophe Ambroise. Analyzing microarray gene expression data. Vol. 422. John Wiley & Sons, 2005.

[14] Arpad, Todor. "Willing to Pay to Save the Planet? Evaluating Support for Increased Spending on Sustainable Development and Environmentally Friendly Policies in Five Countries." PLOS ONE, vol. 13, no. 11, Nov. 2018, p. e0207862. PLoS Journals, doi:10.1371/journal.pone.0207862.

[15] Shonfield, Peter. "LCA of management options for mixed waste plastics." WRAP, UK (2008).

[16] Country data. http://data.stats.gov.cn/. Accessed 17 Feb. 2020.

[17] "Shadow Price." Wikipedia, 11 Dec. 2019. Wikipedia, https://en.wikipedia.org/w/index.php?title=Shadow_pi

[18] Hillier, Frederick S. Introduction to operations research. Tata McGraw-Hill Education, 2012.

[19] "Lagrange Multiplier." Wikipedia, 17 Dec. 2019. Wikipedia, https://en.wikipedia.org/w/index.php?title=Lagrange_multiplier&oldid=931158065.

[20] CGE model-encyclopedia of MBA think tanks. https://wiki.mbalib.com/wiki/CGE. Accessed 18 Feb. 2020.

[21] Analytic Hierarchy Process - Wikipedia. https://en.wikipedia.org/wiki/Analytic_hierarchy_process. Accessed 18 Feb. 2020.

[22] M. M. Li and B. Verma, "Nonlinear curve fitting to stopping power data using rbf neural networks," Expert Systems with Applications, vol. 45, pp. 161–171, 2016.

[23] M. R. Stoline, "The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way anova designs," The American Statistician, vol. 35, no. 3, pp. 134–141, 1981.

[24] D. Tang, B. Qin, T. Liu, and Y. Yang, "User modeling with neural network for review rating prediction," in Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[25] T. Subkhankulova and F. J. Livesey, "Comparative evaluation of linear and exponential amplification techniques for expression profiling at the single-cell level," Genome biology, vol. 7, no. 3, p. R18, 2006.

[26] X. Zhang and C. Dellarocas, "The lord of the ratings: is a movie's fate is influenced by reviews?" ICIS 2006 proceedings, p. 117, 2006.