An Empirical Study on Two-child Policy in China Based on Statistical Analysis and Machine Learning

Yizhou Chi
Shenzhen College of International Education
Shenzhen, China
cyzus@outlook.com

Xingyue Huang
Shenzhen College of International Education
Shenzhen, China
huangxingyue99@hotmail.com

Yu Zhou[*]
Shenzhen University
Shenzhen, China
zhouyu_1022@126.com

Abstract—Since the universal two-child policy (TCP) in China is launched in 2016, many researchers have dedicated their efforts into investigating the influences from the society point of view. In this paper, we look at this issue from a different angle, trying to investigate how the factors influence whether an expectant mother would bore a second child in China empirically. The real-world data from both rural and city regions are used to train an imbalance classification model. In addition, some statistical hypotheses are also made to justify the relevance of these factors. Experimental results demonstrate the validity and effectiveness of our trained model.

Keywords—Two-child policy, Statistical analysis, Imbalance classification, Machine learning, Hypotheses test

I. INTRODUCTION

One-child policy, a population planning policy of China, had been into effect since 1979. The policy aimed at reducing the fertility rate and controlling the population explosion in China. According to the Chinese government, 400 million births were prevented. Although there has been long debates over this aggressive plan, China finally decided to loosen the policy in the end of 2015 as aging population and lack of labor became an increasingly serious issue.

In October 2015, Chinese government announced the plans of abolishing One-child policy, allowing all families to have two children since 2016, in order to address the aging issue. In 2018, after about two years the new policy – two-child policy (TCP) came into effect, it is believed that China is now facing new ramifications from the policy [1].

One concern is gender equality. Long before the One-child policy is established in China, gender inequality existed. Especially in rural part of China where the traditional concept of sexuality plays an important role in people's lives, males are born with pride whereas females are born with shame. Although biologically the chances of a newborn being either male or female solely depends on its chromosome, and such chances are equal, it is reasonable to infer that a considerable amount of expectant mother, after learn their child's sexuality, determine to erase them from existence in the society either by abortion or deliberately letting them hide from any investigation. These would affect the sex ratio of certain area in a large extent [2]-[4]. Also, as the two-child policy is established, the chance of getting a male descendant increases for those expectant mothers. As a result, we aim to investigate the potential relationship between the original sex ratio of rural area against those in urban area. the sex ratio before the

establishment of two-child policy and that after the two-child policy, and the sex ratio of the first child of an expectant mother against that of the second or more child of her [5][6].

Another concern is the adding cost of raising another child. Due to a consistent growth of general price level in RMB, the standard of living for an individual, especially for those in rural area, starts to drop in a terrific speed. Raising another child will no doubt be more expensive than before, and the economic burden raised by having the second child might discourage certain family, even though they are allowed to do so. Assuming the average income from urban area is greater than that from rural area, it is probable to hypothesize that the quantity of new-born second child in urban area will be greater than that in rural area [7-9].

Finally, the ages of pregnancy may be brought forward assuming that expectant mothers had a reasonable schedule of bring their descendants. Since some of those expectant mothers decide to have two or more children instead of one, it can be inferred that they would start their first pregnancy period earlier then before in order to leave enough time behind to have the second child. Such shift in preference of pregnancy age might results in rising demand of Hospital when treating the younger expectant mothers.

In this paper, we have gathered two data sets which record the information of pregnant mothers from two hospitals in the real world application, Hospital A and Hospital B to examine the policy's effects. Hospital A is located relatively in rural area where most of the patients are from villages and small towns while Hospital B is located in urban area where most of the patients are living in the city. We specifically looked up the change in demand for the second-child and that in the sex ratio of newborns. By applying test hypothesis, we were able to confirm a number of statistical significance of changes after the practice of TCP. Also, we applied the machine learning algorithms to design a model for the imbalanced classification to determine whether a newborn is a second child.

II. DATA ANALYSIS

A. Test Hypothesis

Test Hypothesis is a method of statistical inference that is testable on the basis of observing a process through a set of random variables. Usually, a null hypothesis assuming two sets of data have no relationship established, noted as H0. Then, an alternative hypothesis assuming two sets of data have certain relationship is established, noted as H1. The statistic

assumption is made and confirmed valid. A suitable T-statistics is selected and carried out. In this case, because the variance of the population data (i.e. the variance of age of expectant mothers or quantity of second infants born in whole China per month) is unknown, the student's T distribution is selected for the test statistics.

$$t = \frac{\overline{X} - \mu}{\frac{\sigma_x}{\sqrt{n-1}}} \qquad (1)$$

For individual T-test, the degree of freedom must be established as the number of testing sample subtracts 1. Critical region of rejection is then selected based on the types of hypothesis. In our experiments, the significant level, noted as α, is always selected as 5% (0.05). If the t-value calculated from student's T distribution is less than 0.05, we will reject the null hypothesis and reasonably believe that these two factors being have some correlation. Otherwise, we cannot reject the null hypothesis, and we cannot assert that the factors being tested have any relationship. Test hypothesis is a powerful tool to assume features that may contain relationship between any two of them. All the hypotheses in our problem are listed in Fig. 1.

B. Explanations for Hypotheses

1.Effect of TCP on demand
H0: mean of two-child infants before TCP is the same as mean of two-child infants after TCP
H1: the mean number of two-child infants before TCP is less than the mean number of two-child infants after the TCP
2.Relation between age and demand for second children
H0: The correlation coefficient of age and demand equals to 0
H1: The correlation coefficient of age and demand is not equal to 0
3.TCP's effect on age of having the first child
H0: the sex ratio (percentage of male infants) is the same after TCP
H1: the sex ratio is greater from that after TCP
4.TCP's effect on sex ratio of the newborn infants
H0: the sex ratio (percentage of male infants) is the same after TCP
H1: the sex ratio is greater from that after TCP
5.Difference of sex ratio between first and second child
H0: the sex ratio is the same at the second or more child than the first child
H1: The sex ratio is higher at the second or more child than the first child
6.Difference between sex ratio in rural and in city
H0: The sex ratio is the same at the second child (and subsequent children) in the rural area and in the city.
H1: The sex ratio is the higher at the second child (and subsequent children) in the rural area and in the city.
7.Difference emergency ratio for second-child expectant mother and first-child expectant mother
H0: The emergency ratio for second-child mother and first-child mother is equal
H1: The emergency ratio for second-child mother is greater than that of first-child mother

Fig. 1.   The t-value of student's t distribution Hypotheses in our problem

C. Results and Analysis of Test Hypotheses

TABLE I.   STATISTICAL RESULTS FOR THE TEST HYPOTHESES

Hypothesis#	Hospital A			Hospital B		
	t-value	*sample#*	*p-value*	*t-value*	*sample#*	*p-value*
1	-1.5630	15	0.9298	2.9933	25	0.003
2	-1.3120	28	0.2006	0.4688	23	0.6438
3	9.8584	1050	0	-2.8968	2691	0.9981
4	-0.1839	14	0.5715	-0.5747	25	0.7146
5	1.2543	53	0.1077	0.7178	32	0.2391
6	N/A	N/A	N/A	-0.0241	32	0.5096
7	1.0000	8683	0.1587	3.1016	1738	0.001

Given that two datasets are from two different hospitals locating in different regions, the main targeting groups are not the same. In the Hospital A, the patients are mainly from rural area; while the patients in the Hospital B are mainly from urban area. Hence, there are different responses to the enactment of the two-child policy, which are shown in Table I.

The increase in demand for a second child in Hospital B (urban area) is much more significant than that in rural area is possibly due to the historical reason that the one-child limit was most strictly enforced in densely populated urban areas but was relatively loose in rural areas. In some rural regions, a family was allowed to have a second child if the first child was not a boy before the policy was enacted. Therefore, urban areas have experienced a greater increase in demand for a second child.

The third test hypothesis indicates that the Two-child Policy does not have a significant effect in the mothers' age to have the first child in urban areas. However, there is a significant reduction in the mother's average age to have the first child in rural areas. It can be explained as the policy does not urge the urban family to have the first child significantly early because of the pressure people face in the city; meanwhile, the policy might make the family in rural areas willing to have a second child earlier. The fourth test

hypothesis experiment shows that we cannot reject the H0 hypothesis that the sex ratio is the same before and after the establishment of TCP. We assume that Chinese citizens now are less likely to be influenced by the traditional idea of gender or they are less radical on aborting the female infants once detected.

In the fifth and sixth test hypothesis experiments, we cannot reject the H0 hypothesis. This might also suggest an overall reduction in sexual inequality. In addition, such phenomenon may imply the reduction of abortion rate since people are less likely to abort their female infant once discovered. In test hypothesis number six, it is not significant to say the sex ratio is different between the rural area and in the city.

Eventually, in the seventh test hypothesis, two datasets have shown a different conclusion on whether the emergency calling rate is related to whether the mother is having her first child or second child. More research should be conduct to find out the potential relationship.

D. Multi-variable analysis

Despite hypothesis testing, we sorted to develop a model that determines the likeliness of a child that is not the first child using data from Hospital B. Before training, we removed the features that have low correlation coefficients or show insignificant influence based on our results of hypothesis tests. Moreover, some features that have strong correlation coefficients but have low universal significance (e.g. doctors' names) are also screened out. Eventually, we include seven prominent features – age, birth, occupation, marriage, emergency, sex, policy, and location, shown in Table II.

TABLE II.   THE FEATURES CONSIDERED IN THIS PAPER

Feature	Description
Age	Mother's age
Birth	Newborn's birth (month and year)
Occupation	Mother's occupation before childbearing
Marriage	Marriage status (married, single, widowed, other)
Emergency	The extent of emergency (normal, emergent, very emergent)
Sex	Newborn's sexuality
Policy	Whether the two-child policy is enforced
Location	Mother's living location (urban or rural)

E. Weighted Classifier-ensemble

One approach is to create a weighted classifier - a weighted classifier is a combination of classifiers including KNN [10], Decision Tree [11], Support Vector Machines [12] and Naive Bayes [13]. Each classifier will run through the dataset and predict a value; then each value will be weighted and a final prediction will be given.

$$C_{ensemble}(\text{x}) = C_{DT}(x) \times w_{DT} + C_{KNN}(x) \times w_{KNN} + C_{svm}(x) \times w_{svm} + C_{NB}(x) \times w_{NB} \qquad (2)$$

The classifier-ensemble can be efficient since it combines multiple machine-learning algorithms to eventually produce a relatively reliable result minimizing an extreme or occasionally incorrect output produced by one classifier.

F. Boosting and Boostrap Aggregating Algorithms(Bagging)

The Boosting algorithm [14] is adopted based on the weighted classifier in order to improve the result of classification. In the classification process, a weighted system for the dataset is created - each sample will be given a weight. The dataset will be then classified a number of times; after each classification, the sample that is mistakenly classified will gain weight while the sample that is correctly classified will lose weight. Through boosting, the classifier is able to improve its accuracy as it focuses more on the samples that are misclassified.

Bootstrap aggregating algorithm [15] is a method which increases the accuracy and stability of machine learning algorithm. The typical approach is to generate a number of new training sets by sampling from original dataset uniformly and with replacement. Each fitted model from the training sets are then combined by averaging the output. Hence a more accurate model could be given. In this algorithm, our repetition value is set to 9, which is a result of balancing bagging performance and the complexity of the algorithm. An odd number is chosen in order to make sure each sample has an explicit final result. Experiments have shown that utilizing bagging algorithm can significantly improve the accuracy and AUC of classifiers.

G. Solution to Imbalanced Data

During the investigation process, the problem of imbalanced data is found due to the huge number difference between first-child number and second-child number, which has a ratio of 9:1. Using imbalanced data to train the model might lead to the problem of Accuracy Paradox, which states that predictive models with a given level of accuracy may have greater predictive power than models with higher accuracy [16]. To address this problem, the following algorithm is presented.

The strategy of combining random under-sampling and over-sampling is adopted after several attempts to address the problem of imbalance data. Merely randomly choosing and deleting the majority sample of 'first-child' and replicating the minority sample of 'second-child' is sufficient to create a relatively unbiased balanced data-set for training purpose. However, such combining algorithm does have several drawbacks including missing the potentially important concept and overfitting. For future improvement, other imbalanced treating methods such as synthetic minority oversampling techniques (SMOTE) or Cluster-Based Sampling Method could be used to address these problems [17].

H. Result of Multivariable & Analysis

Neglecting specific methods adopted to train the model, the accuracy is relatively stable within the range between 55% – 65% as a mean that the model is quite robust and reasonable. Noticeably, the combination of Decision Tree and Booting algorithm returns the greatest accuracy of 64.7%.

Our classifiers are thoroughly tested with accuracy and the corresponding ROC graphs are plotted in Fig. 2. ROC graphs of the first five classifiers which have the overall best performance are shown in Table III.

Considering the overall performance, Decision Tree classifier is proven to provide the greatest accuracy as well as

the highest AUC, meaning that such classifier has a better performance. In addition, Bagging and Boosting techniques are able to raise overall accuracy and AUC in a significant level. All of AUC value are ranged between 0.70 to 0.77, indicating that all of the classifiers are able to provide reliable prediction as AUC >> 0.5. However, even with these treatments, the accuracy is still under 70%. This shows a lack of deciding features to determine the prediction.

Considering that the features selected are limited because of the restricted information provision by the hospitals, we are unable to collect more important features that are potentially critical to determine the identity of the newborns. Nevertheless, the result is robustly greater than 50% meaning that the trained model with the features included shows a statistically significance in determination.

TABLE III.   THE CLASSIFICATION RESULTS OF DIFFERENT METHODS

| Method	Classifier(s)	Random sampling times	Accuracy
Weighted ensemble	KNN, SVM, Naïve-Bayes, Decision Tree	2000	62.3%
Weighted ensemble + Bagging	KNN, SVM, Naïve-Bayes, Decision Tree	2000	57.7%
Single classifier	KNN	2000	59.8%
Single classifier + Bagging	KNN	2000	60.4%
Single classifier + Bagging	Naïve-Bayes	2000	55.6%
Single classifier + Bagging	Decision Tree	2000	64.2%
Single classifier + Boosting	Decision Tree	2000	64.7%
Single classifier + Bagging + Boosting	Decision Tree	2000	64.8%
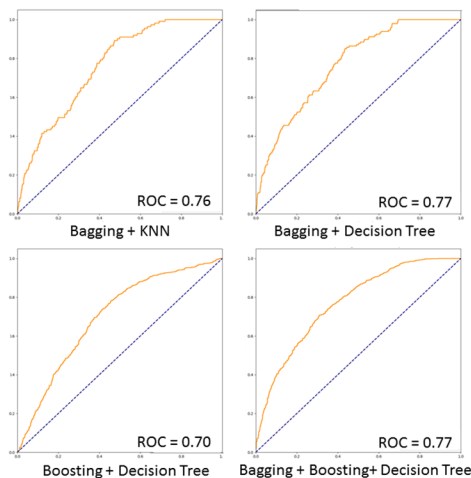
Fig. 2.   ROC curves for different approaches

III.   CONCLUSIONS

One of the goals of this paper was to investigate the connection between several factors such age, sex ratio, etc. We statistically proved that correlation exist in the TCP's effect with second-child demand, second child mother's emergency ratio, and ages to have first kid. Another goal of this paper was to develop a prediction model which might determine whether an infant is a second child or not given a set of features. Our prediction model successfully handles with the problem of imbalanced data and provides us a glimpse of correlation between various factors.

REFERENCES

[1] Zhu, Wei Xing, Li Lu, and Therese Hesketh, "China's excess males, sex selective abortion, and one child policy: analysis of data from 2005 national intercensus survey," Bmj 338 (2009): b1211.
[2] Mei-Dong Q I, Dai M Y, Zheng Y Y, et al, "Discuss the impact and tendency of the Universal Two-child Policy on China's birth rate fluctuation," China Population Resources & Environment, 2016.
[3] Xu B, Pak M. "Gender ratio under China's two-child policy," Journal of Economic Behavior & Organization, vol. 119, pp. 289-307, 2015.
[4] Huang W, Liang J, CNPOPOrg, et al, "Population Problems after the Universal Two- Child Policy," Journal of China Executive Leadership Academy Yanan, 2016.
[5] Wang P, "On the Two-Child Policy in China," China Population Today, 2016(2).
[6] Huang K, "Study on the Impact of Selective Two-child Policy on Sex Ratio at Birth," China Population Today, 2015(6), pp. 48-48.
[7] Sun L J, Sha Y X, Dang Y Q, et al, "Leslie prediction model of Heilongjiang province with "Sparate Two- Child " policy," Journal of Qiqihar University, 2015.
[8] Gong W, Dong X, Caine E D, "Challenges arising from China's two-child policy," Lancet, vol. 387, no. 10025, pp. 1274-1274, 2016.
[9] Li X, Chen X, Fang F, et al, "The "Two-child" Policy and the Premium of School District Housing: Policy Evaluation Analysis Based on the Change of Population Policy," Journal of Finance & Economics, 2017.
[10] Guo G, Wang H, Bell D, et al, KNN Model-Based Approach in Classification[C]// Otm Confederated International Conferences "on the Move To Meaningful Internet Systems, Springer, Berlin, Heidelberg, 2003:986-996.
[11] Liu W, Chawla S, Cieslak D A, et al, A Robust Decision Tree Algorithm for Imbalanced Data Sets[C]// Siam International Conference on Data Mining, SDM 2010, April 29 - May 1, 2010, Columbus, Ohio, Usa. DBLP, 2010:766-777.
[12] Adankon M M, Cheriet M, "Support Vector Machine," Computer Science, vol. 1, no. 4, pp. 1-28, 2009.
[13] Rish I, "An empirical study of the naive Bayes classifier," Journal of Universal Computer Science, vol. 1, no. 2, p. 127, 2001.
[14] Fern A, Givan R, "Online Ensemble Learning: An Empirical Study," Machine Learning, vol. 53, no. 1-2, pp. 71-109, 2003.
[15] Sun Q, Pfahringer B, "Bagging Ensemble Selection,"vol. 7106, pp. 251-260, 2011.
[16] Menon A K, Narasimhan H, Agarwal S, et al, On the statistical consistency of algorithms for binary classification under class imbalance[C], International Conference on Machine Learning, pp. 603-611, 2013.
[17] Haibo, and Edwardo A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge & Data Engineering, 9 (2008), pp. 1263-1284.