Research Article

Research on Application of Naive Bayes Algorithm Based on Attribute Correlation to Unmanned Driving Ethical Dilemma

Guoman Liu,1,2 Yufeng Luo,1 and Jing Sheng2

1College of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China
2School of Mechanical Engineering, Nanchang Institute of Technology, Nanchang 330099, China

Correspondence should be addressed to Guoman Liu; jerry@nit.edu.cn

Received 14 April 2022; Revised 24 June 2022; Accepted 7 July 2022; Published 1 August 2022

At present, unmanned driving technology has made great progress, while those research on its related ethical issues, laws, and traffic regulations are relatively lagging. In particular, it is still a problem how unmanned vehicles make a decision when they encounter ethical dilemmas where traffic collision is inevitable. So it must hinder the application and development of unmanned driving technology. Firstly, 1048575 survey data collected by Moral Machine online experiment platform is analyzed to calculate the prior probability that the straight being protector or sacrificer in ethical dilemmas with single feature. Then, 116 multifeature ethical dilemmas are designed and surveyed. The collected survey data are analyzed to determine decision-making for these ethical dilemmas by adopting the majority principle and to calculate correlation coefficient between attributes, then an improved Naive Bayes algorithm based on attribute correlation (ACNB) is established to solve the problem of unmanned driving decision in multifeature ethical dilemmas. Furthermore, these ethical dilemmas are used to test and verify traditional NB, ADOE, WADOE, CFWNB, and ACNB, respectively. According to the posterior probability that the straight being protector or sacrificer in those ethical dilemmas, classification and decision are made in these ethical dilemmas. Then, the decisions based on these algorithms are compared with human decisions to judge whether these decisions are right. The test results show that ACNB and CFWNB are more consistent with human decisions than other algorithms, and ACNB is more conductive to improve unmanned vehicle’s decision robustness than NB. Therefore, applying ACNB to unmanned vehicles has a good role, which will provide a new research point for unmanned driving ethical decision and a few references for formulating and updating traffic laws and regulations related to unmanned driving technology for traffic regulation authorities.

1. Introduction

At present, unmanned driving technology has achieved good development and application. Many automobile manufacturers, such as Tesla, Uber, and NIO in China, have produced L2 or L3-level autonomous vehicles. However, these researches on unmanned driving ethical decisions and formulation of laws or regulations related to unmanned driving technology are still lagging. In particular, how unmanned vehicles make decisions facing inevitable collisions is still a difficult problem. The reasons may be due to two aspects: First, the pre-allocation scheme of accident risk lacks necessary social consensus. Unlike a traditional driver, who can decide on the spot, unmanned vehicles need to make arrangements before an accident. Second, the relationship between unmanned driving and traditional traffic laws, and regulations needs to be coordinated [1, 2]. For example, whether unmanned vehicle can violate traffic laws and regulations to avoid collision in an emergency, traditional traffic laws and regulations need to be updated and improved. Therefore, the China development research foundation released the report “Future Cornerstone—the Social Role and Ethics of Artificial Intelligence” in 2018, which argued that, “From a humanistic perspective, artificial intelligence(AI) has brought about some fundamental problems that have already happened or will happen to shake the foundation of our society”. “AI is also challenging traditional laws, regulations and social norms”. “Considering the far-reaching impact of AI, the whole society needs to make joint efforts to formulate ethical guidelines and
policies for application” [3]. In addition, it is also hard to determine who will be responsible for an accident for unmanned vehicles. Thus, Tesla’s rights protection event at the 2021 Shanghai Auto Show has impacted on application and market promotion of unmanned driving technology. Therefore, it is urgent to study the unmanned driving ethical decision and formulate laws and regulations related to unmanned driving technology.

The current researches on unmanned driving ethical decision are still in the stage of theoretical exploration, in which influencing factors, implementation principles, and framework of unmanned driving ethical decisions are considered by philosophy, sociology, and law et al. For example, Bonnefon et al. surveyed opinions and thoughts of users or owners of unmanned vehicles by questionnaires. The result shows that users or owners of unmanned vehicles are more willing to protect passengers than pedestrians [4], which may violate the utilitarian principle of protecting more people. While Anja K. Faulhaber et al. also demonstrated that human beings tend to make decisions based on the utilitarian principle in an ethical dilemma by VR. In order to protect more pedestrians on the road, the users of unmanned vehicles would rather sacrifice themselves. Especially, when there are 4–5 pedestrians on the road, more users or owners choose to sacrifice themselves to protect pedestrians [5]. However, it is hard to quantify the utility and consequence of various decision-making. The Massachusetts Institute of Technology (MIT) aims to get a comprehensive understanding of mankind would like unmanned vehicles to make decisions in ethical dilemmas, Moral Machine, an online experimental platform, was set up to carry out questionnaires around the world in 2016. The Moral Machine is designed as a multilingual online “serious game” to gather as much data as possible from all over the world to assess moral preferences by learning how respondents would want autonomous vehicles to solve moral dilemmas in the event of an unavoidable accident. So far, too much decision data in 10 languages have been collected from millions of people in 233 countries and territories, and all survey data in the Moral Machine can be accessed and downloaded at https://goo.gl/JXRrBP. According to the survey result, there are three strong preferences: to protect human beings, to protect more lives, and to protect the younger [6]. However, human unilateral preference cannot solve the problem of ethical decision in moral dilemmas with multi-feature. In addition, Kochupillai et al. thought that applying survey data of Moral Machine directly to unmanned driving ethical decision leads to infringement of some people’s right to life, so it is unfair to discriminate against some people with some external attributes [7]. Each user or owner of the unmanned vehicle has a different requirement for unmanned driving ethical decision, a few designers of unmanned vehicles change the right of choosing unmanned driving ethical decision mode from manufacturers and designers to users or owners, a “moral knob” architecture was proposed for unmanned vehicles, which allowed users or owners to choose ethical decision principle or algorithm according to their requirements [8]. However, the “moral knob” architecture makes some users or owners violate the overall interests of mankind for maximizing personal interest, which can bring up a prisoner’s dilemma. Menon et al. proposed an implementation method of unmanned driving ethical decision, which integrated unmanned driving ethical decision function into safety requirements. So that the ethical requirements can be taken into account in the safety design of unmanned vehicles [9]. However, there is nothing to specify a few mechanisms to implement how unmanned vehicles make ethical decisions. Millar et al. proposed to apply machine ethical decision mechanism to unmanned vehicles and design a general ethical evaluation tool to verify and evaluate the effect of unmanned driving ethical decisions [10]. Pickering et al. also proposed an M2D (Module to Decision) mechanism to predict injury to passengers or pedestrians, then unmanned vehicles can choose decision with injury minimization in an ethical dilemma [11]. However, the mechanism only considers the impact of velocity on injury to passengers or pedestrians but does not consider collision direction and location, which is not perfect and comprehensive. Yueh-hua Wu et al. provided a method of ethical decision optimization based on reverse reinforcement learning. When a decision made by a machine is an action that mankind thinks should not be done, the machine is punished to minimize the occurrence of such action. Otherwise, the action taken by the machine is encouraged to increase occurrence of such action [12]. Maximilian Geisslinger et al. studied how unmanned vehicles make router planning and control decisions in ethical dilemmas from risk ethics theory. The ethical decision is implemented by Bayes theory principle, equality principle, and maximum and minimum principle [13]. However, this method does not specify the weight of each principle in the ethical decisions, how to implement is not clear. Evans et al. proposed an Ethical Valence Theory (EVT). When autonomous vehicles are faced with a moral dilemma, a reinforcement learning algorithm is used to strengthen influencing factors in ethical decisions based on public acceptability and opinions, such as age, social relation, occupation and injury, which will be conducive to maximization of public acceptability as its goal to make decision in moral dilemmas [14]. However, the ethical decision mechanism is only in the theoretical stage, how to evaluate ethical valence is also difficult during implementation. Dietmar and Lucie studied the ethical decision of unmanned vehicles from the perspective of social contract. Contract doctrine is explicitly designed to derive standards acceptable to all rational subjects to eliminate disagreement on thorny ethical issues [15]. However, it is a hard task for unmanned vehicles to distinguish between accident stakeholders and non-stakeholders in moral dilemmas, so as to avoid sacrificing negative rights of nonstakeholders to protect positive rights of stakeholders.

In all, the current research on unmanned driving ethical decision have made some progress and developments, but there have been only a few types of research on influence factor, implementation principle, and theoretical method from philosophy, law, and sociology. The consideration is less on human acceptability and technical feasibility, for which it is hard for unmanned vehicles to make a satisfied
ethics, so the ethical decision. So unmanned driving ethical decisions should be considered and studied from aspects of human acceptability and technical feasibility.

2. Naive Bayes Theory

Naive Bayes algorithm is a classification based on probability density analysis, which predicts the probability of unknown events by known events [14, 15], which is very suitable for classification in the case of a small amount of training set data, simple, and has good performance. However, the traditional Naive Bayes algorithm (NB) requires an assumption of independence among all attributes. The classification principle lies in solving the probability \( p(C_j | x_1, x_2, \ldots, x_n) \) of attribute vector \( X(x_1, x_2, \ldots, x_n) \) being any class \( (C_1, C_2, \ldots, C_j) \), where \( p_j \) represents the probability of the attribute vector \( X \) being class \( C_j \), then \( \max(p_1, p_2, \ldots, p_n) \) is the class which \( X \) is. Therefore, the classification problem can be transformed into solving the posterior probability by using known prior probability [16], as shown in formula.

\[
p(C_j | x_1, x_2, \ldots, x_n) = \frac{p(x_1, x_2, \ldots, x_n | C_j)p(C_j)}{p(x_1, x_2, \ldots, x_n)}
\]

(1)

where \( p(C_j) \) represents the probability that an attribute vector \( X \) is class \( C_j \) in all training sets. \( p(x_1, x_2, \ldots, x_n | C_j) \) represents the prior probability of having joint attributes \( x_1, x_2, \ldots, x_n \) under the condition that it is class \( C_j \), \( p(x_1, x_2, \ldots, x_n) \) represents the probability of having joint attributes \( x_1, x_2, \ldots, x_n \) in the training set. Since classification is judged by the maximum posterior probability, the denominator \( p(x_1, x_2, \ldots, x_n) \) is the same during calculating all posterior probabilities. So the probability is only a comparison between both molecular values. Therefore, the solving process of formula 1 is simplified to the process of seeking maximum molecular value, as shown in formula.

\[
C_{NB} = \arg \max_{C_j \in C} p(x_1, x_2, \ldots, x_n | C_j)p(C_j).
\]

(2)

Because each attribute is required to be independent of each other in NB, there is no dependence among these attributes \( x_1, x_2, \ldots, x_n \) under the same class \( C_j \). Therefore, the solving \( p(x_1, x_2, \ldots, x_n | C_j) \) is transformed into the product of each attribute under the condition \( C_j \), as shown in formula.

\[
p(x_1, x_2, \ldots, x_n | C_j) = \prod p(x_i | C_j).
\]

(3)

However, there is an assumption of independence among all attributes in NB, which affects the accuracy and precision of classification and application in reality. So many experts have provided a few improvements and optimizations for the traditional NB. For example, the TAN algorithm, reduces the dependence of arbitrary attributes in NB by discovering the dependencies between attributes, which is realized by adding the association between attribute pairs [17]. Kononenko proposed a semi-naive Bayes classifier, which combines a few related attributes to form so-called combined attributed nodes [18]. Zhang provided a weighted naive Bayes (WNB) model that assigns different weights to each attribute based on the importance of each attribute [19]. However, each attribute weight is the same for all classes. So Jiang proposed a class-specific attribute weighting naive Bayes (CAWNBN), which discriminatively assigned each attribute a specific weight for each class [20]. Moreover, Zhang also provided a class-specific attribute value weighting for naive Bayes. The approach assigned a specific weight to each attribute value, in which the class-specific attribute value weight matrix is learned by either maximizing the conditional log-likelihood or minimizing the mean squared error [21]. Kohavi combined a naive Bayes classifier with a decision tree, then the instance space was divided by a decision tree, by which a local naive Bayes classifier was established on each leaf node [22]. Zheng and Webb proposed a lazy Bayesian rule (LBR) learning technique by using a lazy learning strategy, which applies the lazy technique to the induction of local naive Bayesian rules [23]. Although the algorithm may improve the classification accuracy, its efficiency is very low. So Yu and Bain [24] proposed an optimized weighted lazy learning naive Bayes classification to improve its efficiency. Furthermore, Webb, et al. [25] proposed the average first-order dependent Bayesian model (AODE), which adopted an average strategy. Each of its sub-model is a SPODE model. All the attributes in the network constructed by the SPODE model depend on the same attribute node, which is also called super-parent node. The posterior probability in AODE is calculated by formula.

\[
P(c | x) \propto \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1, j \neq n}^{n} P(x_j | c, x_i).
\]

(4)

Experimental results show that AODE has better classification accuracy and stability than TAN in most data sets. Especially when data distribution in the training set and testing set is inconsistent, the advantage of AODE is more prominent. However, AODE assumes that all attributes contribute equally to class variables, which is unscientific. So some experts propose a weighted strategy based on attributes (WAODE) for the problem [26]. When calculating the joint probability of attribute node and class variables, a weight is assigned to each attribute to increase the influence on the classification decisions for these important attributes and weaken those unimportant attributes. Weighting strategies have two methods in general, one is weighting based on information gain ratio, and another is weighting based on mutual information. Experiments show that WAODE has a better classification effect, which takes weighting based on mutual information between attributes and class variables [27]. Furthermore, the posterior probability in WAODE is calculated by the following formula.

\[
P(c | x) \propto \frac{1}{n} \sum_{i=1}^{n} w_i P(c, x_i) \prod_{j=1, j \neq n}^{n} P(x_j | c, x_i),
\]

(5)

where \( w_i \) represents the weight coefficient of each attribute, which may be calculated by formula.

\[
w_i = \frac{I(x_i, c)}{\sum_{i=1}^{n} I(x_i, c)}.
\]

(6)
Bayesian classifier is constructed by formula. Proposed a correlation-based feature weighting (CFWNB) weighting model (WAODF). Furthermore, Jiang et al. [29] order dependent Bayesian forest model (AODF) and multiple sub-networks will be established to form a first-order dependent Bayesian forest model (AODF) model. Firstly, the attribute node with a weak node and child node. So Zhong [28] proposed AODF and whether there is a real correlation between super-parent and child node. As the mutation node. Then a node and child node. Therefore, AODE model, which is called mutation node. F_\text{hen a}

\text{WAODF models. Firstly, the attribute node with a weak node and child node. So Zhong [28] proposed AODF and whether there is a real correlation between super-parent and child node. As the mutation node. Then a node and child node. Therefore, AODE model, which is called mutation node. F_\text{hen a}

mathematical problems in engineering

\text{NI(A; C) is the normalized I(A; C) representing mutual relevance between attribute A and class variable C. NI(A; A_j) is the normalized I(A; A_j) representing mutual redundancy between attribute A_i and A_j. They are shown in formulas (9) and (10).}

\text{\text{\text{\text{NI(A; C) = \frac{I(A; C)}{m \sum_{i=1}^{m} I(A; C)}.}}}}

\text{\text{\text{\text{NI(A; A_j) = \frac{I(A; A_j)}{m(m-1) \sum_{i=1}^{m} \sum_{j=1; j\neq i}^{m} I(A; A_j)}}.}}}

\text{Since D_1 may be negative and the weights required by feature weighted naive Bayes should be positive, a standard logistic sigmoid function can be used to transform w_i into the range (0, 1) by formula.}

\text{\text{\text{\text{w_i = \frac{1}{1 + e^{-\text{\frac{D_1}{2}}}}.}}}}

\text{In addition, considering attribute order and the causal relationship between parent and child attributes, hidden naive Bayes (HNB) was proposed by Jiang, in which a hidden parent is created for each attribute that combines these influences from all other attributes [30]. In HBN, the training data set is firstly divided into \text{m} training subsets by different attribute values of the class. For each training subset, the causal relationships between parent and child attributes and attribute order are determined based on the principle of minimum conditional entropy, and sub-modules are constructed. Then a test instance is preassigned with class labels and a pseudo training set is constructed.}

\text{The causal correlation between attributes is determined by the principle of minimum local conditional entropy. Furthermore, a sub-model is constructed. So classification is realized by the \text{m} sub-models [31]. However, this model is constructed by conditional entropy among many attributes, which means only the training set is large enough, that the correlation between attributes can be better determined. Moreover, Jiang et al. proposed a simple, efficient deep feature weighting (DFW), which estimated the conditional probability of naive Bayes by deeply computing feature-weighted frequencies from training data [32]. Considering when the number of the training sets is small and attributes are larger, the training samples are not enough to cover all attributes, the category conditional probability will be equal to 0, and it is impossible to achieve accurate classification. So a few experts proposed a way to solve the problem, in which Laplacian calibration is used [33]. However, the correlation between both attributes needs to be taken into account in ACNB, the p(x_1, x_2, ..., x_n | C_j) is that the original product is multiplied by the maximum correlation coefficient corr(x_i, x_j) between two attributes [34, 35], as shown in formula.

\text{\text{\text{\text{p(x_1, x_2, ..., x_n | C_j) = \text{corr}(x_i, x_j) \prod p(x_i | C_j).}}}}

\text{Since p(C_j) and p(x_i | C_j) are both known prior probabilities, the maximum posterior probability that the feature X is class C_j is calculated to determine which class the feature X belongs to.}

\text{2.1. Prior Probability and Attribute Correlation. In order to get the prior probability that the straight is a protector or sacrificer in an ethical dilemma with single feature, 1048575 data on Moral Machine have been analyzed and studied to calculate prior probabilities of being protector and sacrificer in these ethical dilemmas with single feature, which is shown in Table 1.}

\text{Considering fitness, which is difficult to be recognized and distinguished by machines, the attribute is not considered in this paper. Only six attributes of gender, number, age, social status, traffic rule, and species are regarded as influencing factors of unmanned driving ethical decisions.}

\text{In addition, 116 multi-feature ethical dilemmas are designed and surveyed to achieve human decision data through the online and offline survey. 1006 people were surveyed to collect 18168 data items. According to the survey results, correlation coefficients between attributes are calculated. The decision in ethical dilemmas will be determined by the majority principle, which is determined on the basis of the probability of choosing as protector and sacrificer. For example, an ethical dilemma in Figure 1 An old man is crossing the road on a green light on the straight road. At the same time, a man is also crossing the road on a red light in the turning. When an unmanned vehicle cannot stop immediately, it must choose a pedestrian on one road to collide. According to survey data, 51.6% of respondents choose to turn and 48.4% choose to go straight in the ethical dilemma. So the old man in the straight is considered a protector, the unmanned vehicle should choose to turn to be consistent.
with human choice. The decision in other ethical dilemmas will also be determined in the same way. The correlation coefficient between both attributes can be calculated by the following formula.

\[
\text{Corr}(X_i, X_k) = \frac{P(x_i, x_k | C_i)}{P(x_i | C_i)P(x_k | C_i)} \quad (i, k = 1, 2).
\]

(13)

\(C_i\) is divided into two classes, \(C_1\) is protector, and \(C_2\) is sacrificer. Therefore, attribute correlation should also be considered from two classes respectively. According to attribute values and decision data in 116 ethical dilemmas, the correlation coefficient between attributes may be calculated for protector as shown in Table 2. On the other hand, correlation coefficient may be calculated for sacrificer as shown in Table 3.

Since all respondents choose to protect mankind than others in 116 ethical dilemmas, the correlation between species and other attributes may be ignored.

2.2. Attribute Value and Attribute Vector. In order to study unmanned driving ethical decisions in moral dilemmas, it is necessary to quantify attributes in all ethical dilemmas to get attribute values and attribute vectors in all ethical dilemmas. According to six features (gender, number, age, social status, traffic rule, and species), these attributes in each moral dilemma need to be quantified, and each attribute value is calculated by comparing the attributes of the straight road with that of the turning. For example, when the pedestrian in the straight is female and that in the turning is male, the attribute value is 1 for gender in the straight. Otherwise, when pedestrian in the straight is male and that in the turning is female, the attribute value is 2. When gender is the same, the attribute value is 0. For number, when road users in the straight are more than that in the turning, the attribute value is 1 for the number in the straight. When they are less, the attribute value is 0. When a pedestrian in the straight is younger than that in the turning, the attribute value is 1 for age. When pedestrian in the straight is older, the attribute value is 2. Otherwise, the attribute value is 0 for the same age. In the same way, when the pedestrian’s social status in the straight is higher than that in the turning, the attribute value is 1. Otherwise, the attribute value is 2 for a lower social status. If their social status is the same or unclear, the

<table>
<thead>
<tr>
<th>Feature</th>
<th>Gender</th>
<th>Number</th>
<th>Age</th>
<th>Social status</th>
<th>Traffic rule</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protector</td>
<td>Female</td>
<td>0.509</td>
<td>0.365</td>
<td>0.751</td>
<td>0.197</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.491</td>
<td>0.635</td>
<td>0.249</td>
<td>0.803</td>
<td>0.264</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Gender</th>
<th>Number</th>
<th>Age</th>
<th>Social status</th>
<th>Traffic rule</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protector</td>
<td>Gender</td>
<td>1</td>
<td>2.07</td>
<td>2.026</td>
<td>2.15</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>Number</td>
<td>2.07</td>
<td>1</td>
<td>1.16</td>
<td>1.43</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>2.026</td>
<td>1.16</td>
<td>1</td>
<td>1.65</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>Social status</td>
<td>2.15</td>
<td>1.43</td>
<td>1.65</td>
<td>1</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>Traffic rule</td>
<td>1.89</td>
<td>1.51</td>
<td>1.67</td>
<td>1.60</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Species</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Gender</th>
<th>Number</th>
<th>Age</th>
<th>Social status</th>
<th>Traffic rule</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacrificer</td>
<td>Gender</td>
<td>1</td>
<td>1.5</td>
<td>—</td>
<td>1.56</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>Number</td>
<td>1.5</td>
<td>1</td>
<td>1.29</td>
<td>1.45</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>—</td>
<td>1.29</td>
<td>1</td>
<td>1.42</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>Social status</td>
<td>1.56</td>
<td>1.45</td>
<td>1.42</td>
<td>1</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>Traffic rule</td>
<td>1.88</td>
<td>1.57</td>
<td>1.46</td>
<td>1.56</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Species</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
attribute value is 0. When pedestrians in the straight obey traffic rules and those in the turning violate traffic rules, the attribute value is 1 for traffic rule. If both obey or do not obey traffic rules, the attribute value is 0. Otherwise, the attribute value is 2. Lastly, when the species in the straight is mankind and that in the turning is pet or inanimate, the attribute value is 2 for species. On the contrary, the attribute value is 0. If the species in the straight and the turning is the same, the attribute value is 0. However, an attribute vector is a combination of all attribute values in an ethical dilemma. For the ethical dilemma in Figure 1, Since the differences between the elderly in the straight and the man in the turning are to obey the traffic rule and be older, the corresponding attribute values in the straight are 1 for species. On the contrary, the attribute value is 2. Lastly, when the species in the straight is mankind and that in the turning is pet or inanimate, the attribute value is 2. Lastly, when the species in the straight is mankind and that in the turning is pet or inanimate, the attribute value is 2. Lastly, when the species in the straight is mankind and that in the turning is pet or inanimate, the attribute value is 2. Lastly, when the species in the straight is mankind and that in the turning is pet or inanimate, the attribute value is 2. Lastly, when the species in the straight is mankind and that in the turning is pet or inanimate, the attribute value is 2.

3. Decision Mechanism

According to the above prior probability that the straight being protector or sacrificer, and attribute correlation between attributes, the posterior probability of being protector and sacrificer is calculated in an ethical dilemma by ACNB, which decides the straight will be a protector or sacrificer in an ethical dilemma. The algorithm implementation steps are shown in Table 4.

For example, there is an ethical dilemma in Figure 2.

There are a male doctor and a female doctor in the straight, who obey traffic rules to cross the road. At the same time, a fat lady, two fat men and a thief are crossing the road at a red light. Therefore, the attribute differences between the straight and the turning are number, social status and traffic rule, so the corresponding attribute values are respectively 2, 1, and 1 in the straight. The attribute vector is [0 2 0 1 1 0] in the ethical dilemma. Furthermore, According to survey data on Moral Machine, there are 451304 respondents to survey the straight in all moral dilemmas, 289386 respondents choose the straight as protector, and 161918 respondents choose the straight as sacrificer, so \(p(C_1) = 0.64\) and \(p(C_2) = 0.36\) in formula (1). In addition, there are two attributes: social status and traffic rule, which are conducive to choosing the straight as a protector in the ethical dilemma, so the attribute correlation between social status and traffic rule should be considered during calculating probability. According to Table 2, the correlation coefficient is 1.60. Firstly, the probability of the straight being protector is calculated by formula (12).

\[
p(C_1|x_2, x_3, x_4) = \text{Corr}(x_3, x_4) \frac{p(x_2 = 2, x_3 = 1, x_4 = 1 | C_1) p(C_1)}{p(x_2, x_3, x_4)} \\
= \text{Corr}(x_3, x_4) \frac{p(x_2 = 2 | C_1) p(x_3 = 1 | C_1) p(x_4 = 1 | C_1) (p(C_1))}{p(x_2, x_3, x_4)} \\
= 1.60 \times \frac{0.197 \times 0.683 \times 0.648 \times 0.64}{p(x_2, x_3, x_4)} = \frac{0.0893}{p(x_2, x_3, x_4)}.
\]

On the other hand, the probability of the straight being sacrificer is calculated by formula (1).

\[
p(C_2|x_2, x_3, x_4) = \frac{p(x_2 = 2, x_3 = 1, x_4 = 1 | C_2) p(C_2)}{p(x_2, x_3, x_4)} \\
= \frac{p(x_2 = 2 | C_2) p(x_3 = 1 | C_2) p(x_4 = 1 | C_2) (p(C_2))}{p(x_2, x_3, x_4)} \\
= \frac{0.803 \times 0.317 \times 0.352 \times 0.36}{p(x_2, x_3, x_4)} \\
= \frac{0.0323}{p(x_2, x_3, x_4)}.
\]

\(p(C_1|x_2, x_3, x_4) > p(C_2|x_2, x_3, x_4)\) \(C_{NB} = C_1\).
Therefore, the straight is chosen as a protector in the ethical dilemma by ACNB, unmanned vehicles will choose to turn to protect the pedestrians in the straight. For the human survey on the ethical dilemma, 50.7% of the respondents choose to turn and 40.3% choose to go straight. So the decision of ACNB is consistent with the human decision.

3.1. Test. In order to test and verify the effect of ACNB in the unmanned driving ethical decisions, the 116 ethical dilemmas are chosen as test instances. Then NB, ADOE, WADOE, CFWNB, and ACNB are respectively used to verify these test instances. The weight coefficient wi of each attribute is calculated by formulas (6) and (11) in WADOE and CFWNB, which are shown in Table 5.

The test results of all algorithms are shown in Table 6. Therefore, ACNB and CFWNB are more suitable for human decision than other algorithms. In addition, probability deviations between being protector and sacrificer in 32 test instances are analyzed, as shown in Figure 3 The absolute probability deviations calculated by ACNB are more than those by NB, and their amplitude is larger, which will be conducive to robustness of unmanned driving ethical decision.

### Table 5: Weight coefficient table.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Gender</th>
<th>Number</th>
<th>Age</th>
<th>Social status</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAODE</td>
<td>0.09</td>
<td>0.30</td>
<td>0.11</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>CFWNB</td>
<td>0.13</td>
<td>0.7</td>
<td>0.31</td>
<td>0.5</td>
<td>0.62</td>
</tr>
</tbody>
</table>

### Table 6: Statistical table of test results.

<table>
<thead>
<tr>
<th>Model</th>
<th>NB</th>
<th>ADOE</th>
<th>WADOE</th>
<th>CFWNB</th>
<th>ACNB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>95</td>
<td>97</td>
<td>100</td>
<td>101</td>
<td>101</td>
</tr>
<tr>
<td>Accuracy</td>
<td>82%</td>
<td>83.6%</td>
<td>86.2%</td>
<td>87.1%</td>
<td>87.1%</td>
</tr>
</tbody>
</table>

![Figure 2: Schematic diagram of an ethical dilemma.](image2)

![Figure 3: Schematic diagram of probability deviation comparison.](image3)
4. Conclusions

According to the current situation and existing problems in unmanned driving ethical decision, it is necessary to study unmanned driving ethical decision model and algorithm from the perspective of natural science, which may increase computability and feasibility of ethical decision. Then correlation coefficient between attributes is calculated to establish the ACNB model for autonomous driving ethical decision, which is tested by 116 multi-feature dilemmas. The test results show that ACNB and CFWNB have a better effect than other algorithms and models. In addition, ACNB has stronger robustness than NB. However, the unmanned driving ethical decision is an important issue related to human life, so the accuracy is still far from practical application, for which further research on other influencing factors and algorithms are required in the future. But research and exploration is important to apply a few improved algorithms based on NB theory to unmanned driving ethical decision, which will increase the computability and feasibility of decision-making in an ethical dilemma for unmanned vehicles. So unmanned vehicle manufacturers may set up some algorithms based on human decision data in unmanned vehicles to realize self-decision in an ethical dilemma. However, these algorithms and models may have some discrimination in some external attributes, just like mankind, resulting in unfair decision of unmanned vehicles. Therefore, when unmanned vehicle manufacturers embed these algorithms into unmanned vehicles, they should set ethical principles in unmanned vehicles in order to not violate human ethics. On the other hand, traffic laws and regulations related to unmanned driving can be formulated and designed by referring to the weight of each attribute, which is conducive to the formulation and improvement of traffic laws and regulations for traffic regulation authorities.

Data Availability

The data used to support the findings of this study are available from the corresponding open platform and previously reported studies and datasets.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Science and Technology Research Project of Jiangxi Education Department (Grant no. GJJ201911), the National Natural Science Foundation of China (Grant no. 51865031).

References


