



(12) 发明专利申请

(10) 申请公布号 CN 118032374 A

(43) 申请公布日 2024. 05. 14

(21) 申请号 202410336952.4

(22) 申请日 2024.03.22

(71) 申请人 清华大学

地址 100084 北京市海淀区清华园

(72) 发明人 封硕 李述 张毅 杨敬轩

何泓霖

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

专利代理师 李丹 栗若木

(51) Int. Cl.

G01M 17/007 (2006.01)

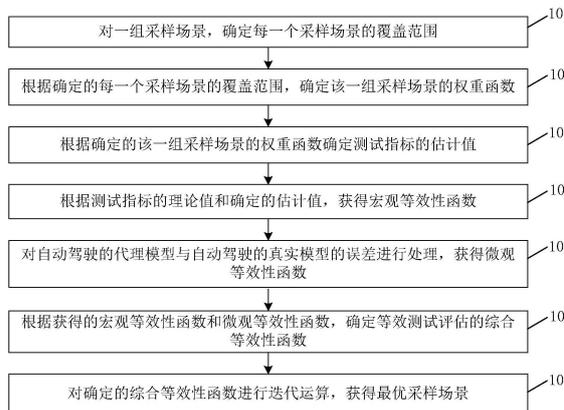
权利要求书2页 说明书8页 附图3页

(54) 发明名称

一种测试评估的方法、计算机存储介质及终端

(57) 摘要

本申请公开一种测试评估的方法、计算机存储介质及终端,包括:对一组采样场景,确定每一个采样场景的覆盖范围;根据每一个采样场景的覆盖范围,确定该一组采样场景的权重函数;根据确定的该一组采样场景的权重函数确定测试指标的估计值;根据测试指标的理论值和估计值,获得宏观等效性函数;对自动驾驶的代理模型与自动驾驶的真实模型的误差进行处理,获得微观等效性函数;根据宏观等效性函数和微观等效性函数,确定等效测试评估的综合等效性函数;对综合等效性函数进行迭代运算,获得最优采样场景。本公开实施例通过综合等效性函数,通过衡量场景覆盖范围生成少量代表性测试的采样场景,保证了少量测试成本下的自动驾驶安全性测试的可靠性。



1. 一种测试评估的方法,其特征在于,包括:

对一组采样场景,确定每一个采样场景的覆盖范围;

根据确定的每一个采样场景的覆盖范围,确定该一组采样场景的权重函数;

根据确定的该一组采样场景的权重函数确定测试指标的估计值;根据测试指标的理论值和确定的估计值,获得宏观等效性函数;

对自动驾驶的代理模型与自动驾驶的真实模型的误差进行处理,获得微观等效性函数;

根据获得的宏观等效性函数和微观等效性函数,确定等效测试评估的综合等效性函数;

对确定的综合等效性函数进行迭代运算,获得最优采样场景;

其中,所述一组采样场景包含n个所述采样场景,n小于或等于预设数值;所述微观等效性函数表示采样场景 $x_i$ 与其覆盖范围内的其它采样场景的测试指标的差异, $x_i$ 表示n个采样场景中的第i个;所述宏观等效性函数表示通过所有采样场景得到的自动驾驶整体测试指标估计值与理论值的差异。

2. 根据权利要求1所述的方法,其特征在于,所述根据确定的该一组采样场景的权重函数确定测试指标的估计值,包括:

根据所述权重函数,采用加权和的方式计算所述测试指标的估计值。

3. 根据权利要求2所述的方法,其特征在于,所述测试指标的估计值为 $\hat{\mu}$ ,所述估计值的表达式为:

$$\hat{\mu} = \sum_{i=1}^n \mathbb{P}(A | x_i) w(x_i);$$

其中,A表示兴趣事件, $\mathbb{P}(A | x_i)$ 表示所述采样场景 $x_i$ 的测试指标值, $w(x_i)$ 表示所述权重函数满足 $w(x_i) = \sum_{x \in C(x_i)} p(x)$ , $C(x_i)$ 表示采样场景 $x_i$ 的覆盖范围。

4. 根据权利要求1至3任一项所述的方法,其特征在于,所述宏观等效性函数为 $E_1$ ,所述宏观等效性函数的表达式为:

$$E_1 = -|\hat{\mu} - \mu|;$$

式中, $\mu$ 为所述测试指标的理论值, $\hat{\mu}$ 为所述测试指标的估计值。

5. 根据权利要求4所述的方法,其特征在于,所述微观等效性函数为 $E_2$ ,所述微观等效性函数的表达式为:

$$E_2 = -\sum_{i=1}^n \sum_{x \in C(x_i)} [\mathbb{P}(A | x) - \mathbb{P}(A | x_i)]^2 p(x);$$

式中, $C(x_i)$ 表示采样场景 $x_i$ 的覆盖范围, $\mathbb{P}(A | x) - \mathbb{P}(A | x_i)$ 表示对自动驾驶的所述代理模型与自动驾驶的所述真实模型的误差的估计, $p(x)$ 表示所述场景x在自然分布中出现的概率。

6. 根据权利要求5所述的方法,其特征在于,所述采样场景 $x_i$ 的覆盖范围 $C(x_i)$ 的表达式为:

$$C(x_i) = \left\{ x \mid \operatorname{argmin}_i d(x, x_i) = i \right\};$$

其中,  $d(x, x_i)$  为状态空间场景与采样场景距离的距离函数,  $x$  表示状态空间场景中的采样场景。

7. 根据权利要求5所述的方法, 其特征在于, 所述综合等效性函数为  $J(x_1, \dots, x_n)$ , 所述综合等效性函数的表达式为:

$$J(x_1, \dots, x_n) = w_1 E_1 + w_2 E_2;$$

其中,  $w_1$  为所述宏观等效性函数的权重、 $w_2$  为所述微观等效性函数的权重。

8. 根据权利要求1所述的方法, 其特征在于, 所述对确定的综合等效性函数进行迭代运算, 以获得最优采样场景, 包括:

利用梯度下降法迭代所述综合等效性函数, 确定一组采样场景  $x_1, \dots, x_n$ ;

根据确定的采样场景  $x_1, \dots, x_n$  的覆盖范围得到综合等效性函数;

通过最大化所述根据确定的采样场景  $x_1, \dots, x_n$  的覆盖范围得到的综合等效性函数得到最优采样场景。

9. 一种计算机存储介质, 所述计算机存储介质中存储有计算机程序, 所述计算机程序被处理器执行时实现如权利要求1至8中任一项所述的测试评估的方法。

10. 一种终端, 包括: 存储器和处理器, 所述存储器中保存有计算机程序; 其中, 处理器被配置为执行存储器中的计算机程序;

所述计算机程序被所述处理器执行时实现如权利要求1至8中任一项所述的测试评估的方法。

## 一种测试评估的方法、计算机存储介质及终端

### 技术领域

[0001] 本申请涉及但不限于自动驾驶技术,其中涉及一种测试评估的方法、计算机存储介质及终端。

### 背景技术

[0002] 当前国内外自动驾驶汽车的试点性运行面临着显著的安全性问题,阻碍了自动驾驶技术的规模化落地。国外一些公司自动驾驶开放道路运行接连出现事故,导致运营执照遭到吊销、公司面临困难;因此,可靠的安全性测试是自动驾驶投入应用之前的必要环节。开放道路测试是常用的安全性测试方法,由于真实世界中存在大量低风险场景,开放道路测试难以有效测量自动驾驶在危险场景下的表现,测试效率低下;此外,在真实道路执行测试可能为社会带来安全风险。

[0003] 自动驾驶研发迭代周期短,算法更新迅速,难以进行大规模测试;政府、第三方测试机构同样难以在多种自动驾驶算法上进行大规模测试。使用相关技术中的方法,在少量测试次数的限制下进行测试的结果不可靠;基于专家经验预先生成含有关键场景的场景库,选择少量场景进行测试可评估自动驾驶在危险情况下的表现,但是由于需要依赖经验,该方法难以量化测试结果,或测试结果不具备理论上的可靠性保证。

[0004] 综上,如何从理论上保证少量测试成本下的自动驾驶安全性测试可靠性,成为一个有待解决的问题。

### 发明内容

[0005] 以下是对本申请详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

[0006] 本公开实施例提供一种测试评估的方法、计算机存储介质及终端,能够保证少量测试成本下的自动驾驶安全性测试的可靠性。

[0007] 本公开实施例提供了一种测试评估的方法,包括:

[0008] 对一组采样场景,确定每一个采样场景的覆盖范围;

[0009] 根据确定的每一个采样场景的覆盖范围,确定该一组采样场景的权重函数;

[0010] 根据确定的该一组采样场景的权重函数确定测试指标的估计值;

[0011] 根据测试指标的理论值和确定的估计值,获得宏观等效性函数;

[0012] 对自动驾驶的代理模型与自动驾驶的真实模型的误差进行处理,获得微观等效性函数;

[0013] 根据获得的宏观等效性函数和微观等效性函数,确定等效测试评估的综合等效性函数;

[0014] 对确定的综合等效性函数进行迭代运算,获得最优采样场景;

[0015] 其中,所述一组采样场景包含 $n$ 个所述采样场景, $n$ 小于或等于预设数值;所述微观等效性函数表示采样场景 $x_i$ 与其覆盖范围内的其它采样场景的测试指标的差异, $x_i$ 表示 $n$ 个

采样场景中的第*i*个;所述宏观等效性函数表示通过所有采样场景得到的自动驾驶整体测试指标估计值与理论值的差异。

[0016] 另一方面,本公开实施例还提供一种计算机存储介质,所述计算机存储介质中存储有计算机程序,所述计算机程序被处理器执行时实现上述测试评估的方法。

[0017] 再一方面,本公开实施例还提供一种终端,包括:存储器和处理器,所述存储器中保存有计算机程序;其中,

[0018] 处理器被配置为执行存储器中的计算机程序;

[0019] 所述计算机程序被所述处理器执行时实现如上述测试评估的方法。

[0020] 与相关技术相比,本公开实施例通过宏观等效性函数和微观等效性函数确定等效测试评估的综合等效性函数,通过衡量场景覆盖范围生成少量代表性测试的采样场景,保证了少量测试成本下的自动驾驶安全性测试的可靠性。

[0021] 本申请的其它特征和优点将在随后的说明书中阐述,并且,部分地从说明书中变得显而易见,或者通过实施本申请而了解。本申请的其他优点可通过在说明书以及附图中所描述的方案来实现和获得。

## 附图说明

[0022] 附图用来提供对本申请技术方案的理解,并且构成说明书的一部分,与本申请的实施例一起用于解释本申请的技术方案,并不构成对本申请技术方案的限制。

[0023] 图1为本公开实施例测试评估的方法的流程图;

[0024] 图2为本公开实施例自动驾驶汽车的切车场景的示意图;

[0025] 图3为本公开实施例切车场景的自然概率分布示意图;

[0026] 图4为本公开实施例一采样场景选取与覆盖范围估计的效果图;

[0027] 图5为本公开实施例另一采样场景选取与覆盖范围估计的效果图。

## 具体实施方式

[0028] 本申请描述了多个实施例,但是该描述是示例性的,而不是限制性的,并且对于本领域的普通技术人员来说显而易见的是,在本申请所描述的实施例包含的范围内可以有更多的实施例和实现方案。尽管在附图中示出了许多可能的特征组合,并在具体实施方式中进行了讨论,但是所公开的特征的许多其它组合方式也是可能的。除非特意加以限制的情况以外,任何实施例的任何特征或元件可以与任何其它实施例中的任何其他特征或元件结合使用,或可以替代任何其它实施例中的任何其他特征或元件。

[0029] 本申请包括并设想了与本领域普通技术人员已知的特征和元件的组合。本申请已经公开的实施例、特征和元件也可以与任何常规特征或元件组合,以形成由权利要求限定的独特的发明方案。任何实施例的任何特征或元件也可以与来自其它发明方案的特征或元件组合,以形成另一个由权利要求限定的独特的发明方案。因此,应当理解,在本申请中示出和/或讨论的任何特征可以单独地或以任何适当的组合来实现。因此,除了根据所附权利要求及其等同替换所做的限制以外,实施例不受其它限制。此外,可以在所附权利要求的保护范围内进行各种修改和改变。

[0030] 此外,在描述具有代表性的实施例时,说明书可能已经将方法和/或过程呈现为特

定的步骤序列。然而,在该方法或过程不依赖于本文所述步骤的特定顺序的程度上,该方法或过程不应限于所述的特定顺序的步骤。如本领域普通技术人员将理解的,其它的步骤顺序也是可能的。因此,说明书中阐述的步骤的特定顺序不应被解释为对权利要求的限制。此外,针对该方法和/或过程的权利要求不应限于按照所写顺序执行它们的步骤,本领域技术人员可以容易地理解,这些顺序可以变化,并且仍然保持在本申请实施例的精神和范围内。

[0031] 图1为本公开实施例测试评估的方法的流程图,如图1所示,包括:

[0032] 步骤101、对一组采样场景,确定每一个采样场景的覆盖范围;本公开实施例初始化时,可以通过随机采样、拟蒙特卡罗采样(RQMC)等方法生成一组初始采样场景。

[0033] 在一种示例性实例中,步骤101之前,本公开实施例方法还包括:进行场景空间建模,确定采样场景空间、采样场景在自然分布中的概率,测试指标理论值;其中,确定场景空间基本建模,包括:采样场景空间中的场景为 $x \in X$ ,场景在自然分布中出现概率为 $p(x)$ ,测试指标理论值为 $\mu = \sum_{x \in X} \mathbb{P}(A|x)p(x)$ ,其中A表示兴趣事件, $\mathbb{P}(A|x)$ 表示场景空间内任一场景x的测试指标值(如兴趣事件A在场景x中发生的概率)。

[0034] 步骤102、根据确定的每一个采样场景的覆盖范围,确定该一组采样场景的权重函数;

[0035] 步骤103、根据确定的该一组采样场景的权重函数确定测试指标的估计值;

[0036] 步骤104、根据测试指标的理论值和确定的估计值,获得宏观等效性函数;

[0037] 步骤105、对自动驾驶的代理模型与自动驾驶的真实模型的误差进行处理,获得微观等效性函数;本公开实施例中的真实模型为待测实际自动驾驶模型,在每个场景x有测试指标 $\mathbb{P}^*(A|x)$ ,自动驾驶的代理模型是对真实自动驾驶模型的一个估计,在每个场景x有测试指标 $\mathbb{P}(A|x)$ ,以下理论推导和分析中由于实际自动驾驶测试指标 $\mathbb{P}^*(A|x)$ 在测试前未知,均使用 $\mathbb{P}(A|x)$ 作为其估计。

[0038] 步骤106、根据获得的宏观等效性函数和微观等效性函数,确定等效测试评估的综合等效性函数;

[0039] 步骤107、对确定的目标函数进行迭代运算,获得最优采样场景;

[0040] 其中,一组采样场景包含n个采样场景,n小于或等于预设数值;微观等效性函数表示采样场景 $x_i$ 与其覆盖范围内的其它采样场景的测试指标的差异, $x_i$ 表示n个采样场景中的第i个;宏观等效性函数表示通过所有采样场景得到的自动驾驶整体测试指标估计值与理论值的差异。

[0041] 本公开实施例通过宏观等效性函数和微观等效性函数确定等效测试评估的综合等效性函数,通过衡量场景覆盖范围生成少量代表性测试的采样场景,通过少量代表性场景上的测试得到与充分测试等效的评估结果,保证了少量测试成本下的自动驾驶安全性测试的可靠性。

[0042] 在一种示例性实例中,本公开实施例根据确定的该一组采样场景的权重函数确定测试指标的估计值,包括:

[0043] 根据权重函数,采用加权和的方式计算测试指标的估计值。

[0044] 在一种示例性实例中,本公开实施例权重函数也可以采用其他方式确定,例如:常

数权重函数,依据先验知识的重要性权重函数,数据驱动的可学习权重函数;

[0045] 在一种示例性实例中,本公开实施例估计值也可以采用相关技术中的其他方式确定,例如:通用形式估计函数 $f[\mathbb{P}(A|x_i), i=1, \dots, n]$ ;本公开实施例对此不作限制。

[0046] 在一种示例性实例中,本公开实施例测试指标的估计值为 $\hat{\mu}$ ,估计值的表达式可以为:

$$[0047] \quad \hat{\mu} = \sum_{i=1}^n \mathbb{P}(A|x_i)w(x_i);$$

[0048] 其中,A表示兴趣事件, $\mathbb{P}(A|x_i)$ 表示采样场景 $x_i$ 的测试指标值, $w(x_i)$ 表示权重函数满足 $w(x_i) = \sum_{x \in C(x_i)} p(x)$ , $C(x_i)$ 表示采样场景 $x_i$ 的覆盖范围。

[0049] 在一种示例性实例中,本公开实施例宏观等效性函数为 $E_1$ ,宏观等效性函数的表达式为:

$$[0050] \quad E_1 = -|\hat{\mu} - \mu|;$$

[0051] 式中, $\mu$ 为测试指标的理论值, $\hat{\mu}$ 为测试指标的估计值。

[0052] 本公开实施例 $E_1$ 根据测试指标的估计值与理论值反应测试场景的宏观等效性,即能否通过一组小规模测试精确等效被测指标真值。

[0053] 在一种示例性实例中,本公开实施例记自动驾驶汽车的测试指标(例如:事故率)的理论值为 $\mu$ , $\mu = \mathbb{P}(A) = \sum_{x \in X} P(A|x)p(x)$ ,其中,A表示兴趣事件(例如:碰撞事故), $p(x)$ 表示测试场景 $x$ 的自然概率分布, $P(A|x)$ 为兴趣事件A的在采样场景 $x$ 上的测试指标(如事件A在采样场景 $x$ 中的出现概率);

[0054] 相关技术中的测试方法基于蒙特卡洛采样原理实现测试评估,该方法在事件发生概率极低的情况下测试效率低下,因此在实际应用中难以得到有效测试结果。

[0055] 在一种示例性实例中,本公开实施例微观等效性函数为 $E_2$ ,微观等效性函数的表达式为:

$$[0056] \quad E_2 = -\sum_{i=1}^n \sum_{x \in C(x_i)} [\mathbb{P}(A|x) - \mathbb{P}(A|x_i)]^2 p(x);$$

[0057] 式中, $C(x_i)$ 表示采样场景 $x_i$ 的覆盖范围, $\mathbb{P}(A|x) - \mathbb{P}(A|x_i)$ 表示对自动驾驶的代理模型与自动驾驶的真实模型的误差的估计,即用采样场景 $x_i$ 的测试指标值 $\mathbb{P}(A|x_i)$ 与其覆盖范围内的其他场景 $x$ 的测试指标值 $\mathbb{P}(A|x)$ 的差异估计代理模型与真实模型的误差, $p(x)$ 表示场景 $x$ 在自然分布中出现的概率。

[0058] 在一种示例性实例中,本公开实施例采样场景 $x_i$ 的覆盖范围 $C(x_i)$ 的表达式为:

$$[0059] \quad C(x_i) = \left\{ x \mid \underset{i}{\operatorname{argmin}} d(x, x_i) = i \right\};$$

[0060] 其中, $d(x, x_i)$ 为状态空间场景与采样场景距离的距离函数,可以用欧式距离表示, $x$ 表示状态空间场景中的采样场景。

[0061] 在一种示例性实例中,本公开实施例综合等效性函数为 $J(x_1, \dots, x_n)$ ,综合等效性

的表达式为：

$$[0062] \quad J(x_1, \dots, x_n) = w_1 E_1 + w_2 E_2;$$

[0063] 其中,  $w_1$  为宏观等效性函数的权重、 $w_2$  为微观等效性函数的权重。

[0064] 本公开实施例基于专家经验或历史数据确定自动驾驶的代理模型的可靠性,若代理模型可靠性高,设置  $w_1$  增大,  $w_2$  减小,反之,若代理模型可靠性低,设置  $w_1$  减小,  $w_2$  增大。

[0065] 在一种示例性实例中,本公开实施例对确定的综合等效性函数进行迭代运算,以获得最优采样场景,包括:

[0066] 利用梯度下降法迭代综合等效性,确定一组采样场景  $x_1, \dots, x_n$ ;

[0067] 根据确定的采样场景  $x_1, \dots, x_n$  的覆盖范围得到综合等效性;

[0068] 通过最大化根据确定的采样场景  $x_1, \dots, x_n$  的覆盖范围得到的综合等效性函数得到最优采样场景。

[0069] 需要说明的是,本公开实施例可以采用相关技术已有的运算方法执行上述迭代运算,本公开实施例对此不作限制。

[0070] 本公开实施例还提供一种计算机存储介质,计算机存储介质中存储有计算机程序,计算机程序被处理器执行时实现上述测试评估的方法。

[0071] 本公开实施例还提供一种终端,包括:存储器和处理器,存储器中保存有计算机程序;

[0072] 其中,

[0073] 处理器被配置为执行存储器中的计算机程序;

[0074] 计算机程序被处理器执行时实现如上述测试评估的方法。

[0075] 以下通过应用示例对本公开实施例进行简要说明,应用示例仅用于陈述本公开实施例,并不用于限定本公开实施例的保护范围。

[0076] 应用示例

[0077] 本公开实施例提出了一种测试评估方法,是面向自动驾驶的等效测试评估方法,该方法针对实际应用中测试成本限制严格的情况,依据场景空间相似性衡量测试场景的覆盖范围,从而选择少量具有突出代表性的测试场景,并利用等效测试算法将少量测试结果恢复为与充分测试等效的结果,从理论上保证测试成本严格情况下的测试结果可靠性、量化自动驾驶性能指标,在提升测试效率的同时保证测试结果理论上的准确性。

[0078] 本公开实施例记自动驾驶汽车的测试指标(例如:事故率)的理论值为  $\mu$ ,  $\mu = \mathbb{P}(A) = \sum_{x \in X} \mathbb{P}(A|x)p(x)$ , 其中,  $A$  表示兴趣事件(例如:碰撞事故),  $p(x)$  表示测试场景  $x$  的自然概率分布,  $\mathbb{P}(A|x)$  为兴趣事件  $A$  的在场景  $x$  上的测试指标(如事件  $A$  在场景  $x$  中的出现概率);

[0079] 相关技术中的测试评估方法基于蒙特卡洛采样原理,测试指标可以通过以下函数估算:

$$[0080] \quad \hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(A|x_i), x_i \sim p;$$

[0081] 该方法在事件  $A$  发生概率极低时测试效率低下,因此在实际应用中难以得到有效测试结果。

[0082] 开放道路测试使用的方法是蒙特卡洛采样法,由于自动驾驶汽车的事件事件发生概率极低,基于蒙特卡洛采样的测试方法需要极多测试次数才可以得到事故率的估计值,而在测试资源(时间和成本等资源)受限的情况下估计值方差极大,估计结果不可靠。

[0083] 本公开实施例测试评估方法预先确定测试集大小 $n$ ,在测试次数(由于时间、成本等测试资源受限,因此导致测试次数受限)小于预设设定的次数阈值的情况下, $n$ 通常较小(例如 $n=10$ )。本公开实施例将蒙特卡洛采样法扩展至更一般的情形,以预先策略选择采样场景(也可以称作采样点) $x_1, \dots, x_n$ ,并通过加权和计算测试指标的估计值 $\hat{\mu}$ ;

$$[0084] \quad \hat{\mu} = \sum_{i=1}^n \mathbb{P}(A | x_i) w(x_i);$$

[0085] 根据最小化采样场景(采样场景的个数小于或等于预设次数阈值,例如10)的测试误差,迭代选择采样场景;

[0086]  $E_1 = -|\hat{\mu} - \mu|$ 其中, $E_1$ 根据测试指标的估计值与理论值反应测试场景的宏观等效性,即能否通过一组小规模测试精确等效被测指标真值。

[0087] 本公开实施例等效测试评估方法通过场景覆盖范围衡量测试场景代表性,并依据场景覆盖范围赋权重,根据状态空间场景到采样场景的距离表示场景覆盖范围,采样场景 $x_i$ 的覆盖范围 $C(x_i)$ 为:

$$[0088] \quad C(x_i) = \left\{ x \mid \operatorname{argmin}_i d(x, x_i) = i \right\};$$

[0089] 其中, $d(x, x_i)$ 为状态空间场景与采样场景距离的距离函数,可使用欧式距离表示, $x$ 表示状态空间场景中的点;依据采样场景覆盖范围可计算该采样场景的权重函数 $w(x_i) = \sum_{x \in C(x_i)} p(x)$ 进而计算测试指标的估计值 $\hat{\mu}$ 。

[0090] 由于在测试前被测对象未知,而上述优化过程需要估计测试误差,这一过程在自动驾驶的代理模型(Surrogate Model)上实现,而当自动驾驶的代理模型与真实模型之间存在差异时,测试准确性可能受到影响;本公开实施例使用微观等效性函数应对自动驾驶的代理模型与真实模型的误差进行估计:

$$[0091] \quad E_2 = -\sum_{i=1}^n \sum_{x \in C(x_i)} [\mathbb{P}(A | x) - \mathbb{P}(A | x_i)]^2 p(x);$$

[0092] 其含义为每一个采样场景 $x_i$ 与其覆盖范围内的其它采样场景的测试指标的差异,若采样场景与其覆盖范围内的其它采样场景的测试指标相似,则该采样场景的代理模型可靠性可能更强;本公开实施例通过提高微观等效性,最大程度上消除代理模型不准确为测试结果带来的误差。

[0093] 最后,测试评估方法的综合等效性函数为宏观等效性函数和微观等效性函数的结合:

[0094]  $J(x_1, \dots, x_n) = w_1 E_1 + w_2 E_2$ ;其中, $w_1$ 为宏观等效性权重、 $w_2$ 为微观等效性权重;本公开实施例基于专家经验或历史数据确定自动驾驶的代理模型的可靠性,若代理模型可靠性高,设置 $w_1$ 增大, $w_2$ 减小,反之,若代理模型可靠性低,设置 $w_1$ 减小, $w_2$ 增大。通过最优化目标函数,可利用梯度下降法迭代寻找一组采样场景 $x_1, \dots, x_n$ ,计算覆盖范围并进一步得到等

效性函数,通过最大化等效性得到最优采样场景,则这组场景可在保证误差最小的情况下兼顾被测模型先验信息不准确的影响,实现少量场景可靠测试。

[0095] 本公开实施例以最小化误差作为优化的引导,在理论上量化并确保了测试结果的准确性;相较于一般场景,选择代表性场景,提高了自动驾驶的可靠性测试的测试效率,可在少量测试情况下保证测试可靠性。

[0096] 本公开实施例模型覆盖范围 $C(x_i)$ 衡量部分可能使用多种不同形式,也可对距离函数 $d(x, x_i)$ 形式进行调整。

[0097] 以下以图2所示的自动驾驶汽车的切车场景为例,其中被测自动驾驶车辆记为AV,背景车辆记为BV,BV换道后有可能与AV发生碰撞;切车场景定义为 $X := (R, \dot{R})$ ,其中,R表示相对距离(Range), $\dot{R}$ 表示相对速度(Range rate);切车场景的自然概率分布 $p$ 如图3所示;

[0098] 本公开实施例测试评估方法进行测试次数为10时的采样场景选取,图4与图5为本公开实施例采样场景选取与覆盖范围估计的效果图,通过效果图可以看出本公开实施例测试评估方法生成测试场景在场景空间各个部分均有分布,且在一般低危险性区域场景采样数量较少,单个场景覆盖范围较大;在高危险性区域采样场景较多,单个场景覆盖范围较小;因此测试评估方法可充分利用场景空间信息,在测试成本受限严格时保证测试可靠性。

[0099] 根据本公开实施例所提供的测试评估方法对自动驾驶汽车进行测试评估,与蒙特卡洛方法(MC)和基于均匀采样的随机拟蒙特卡洛方法(RQMC)相比,测试指标的估计结果如表1所示;表1中显示了测试次数为10时三种方法在2种不同自动驾驶模型AV-1,AV-2上的测试效果。

方法	平均误差 ( $\times 10^{-3}$ )		方差 $\times 10^{-5}$	
	AV-1	AV-2	AV-1	AV-2
[0100] 蒙特卡洛采样法	1.90	9.59	10.5	40.1
随机拟蒙特卡洛采样法	0.98	4.33	0.086	2.19
等效测试方法	0.97	2.40	0.023	2.12

[0101] 表1

[0102] 通过对比可以看到等效测试方法估计误差与方差在测试次数受限严格时明显小于两种基准方法,提高了测试效率,保证了少量测试结果可靠性。

[0103] 本领域普通技术人员可以理解,上文中所公开方法中的全部或某些步骤、系统、装置中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。在硬件实施方式中,在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分;例如,一个物理组件可以具有多个功能,或者一个功能或步骤可以由若干物理组件合作执行。某些组件或所有组件可以被实施为由处理器,如数字信号处理器或微处理器执行的软件,或者被实施为硬件,或者被实施为集成电路,如专用集成电路。这样的软件可以分布在计算机可读介质上,计算机可读介质可以包括计算机存储介质(或非暂时性介质)和通信介质(或暂时性介质)。如本领域普通技术人员公知的,术语计算机存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、程序模块或其他数据)的任何方法或技术中实施的易失性和非

易失性、可移除和不可移除介质。计算机存储介质包括但不限于RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外,本领域普通技术人员公知的是,通信介质通常包含计算机可读指令、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据,并且可包括任何信息递送介质。

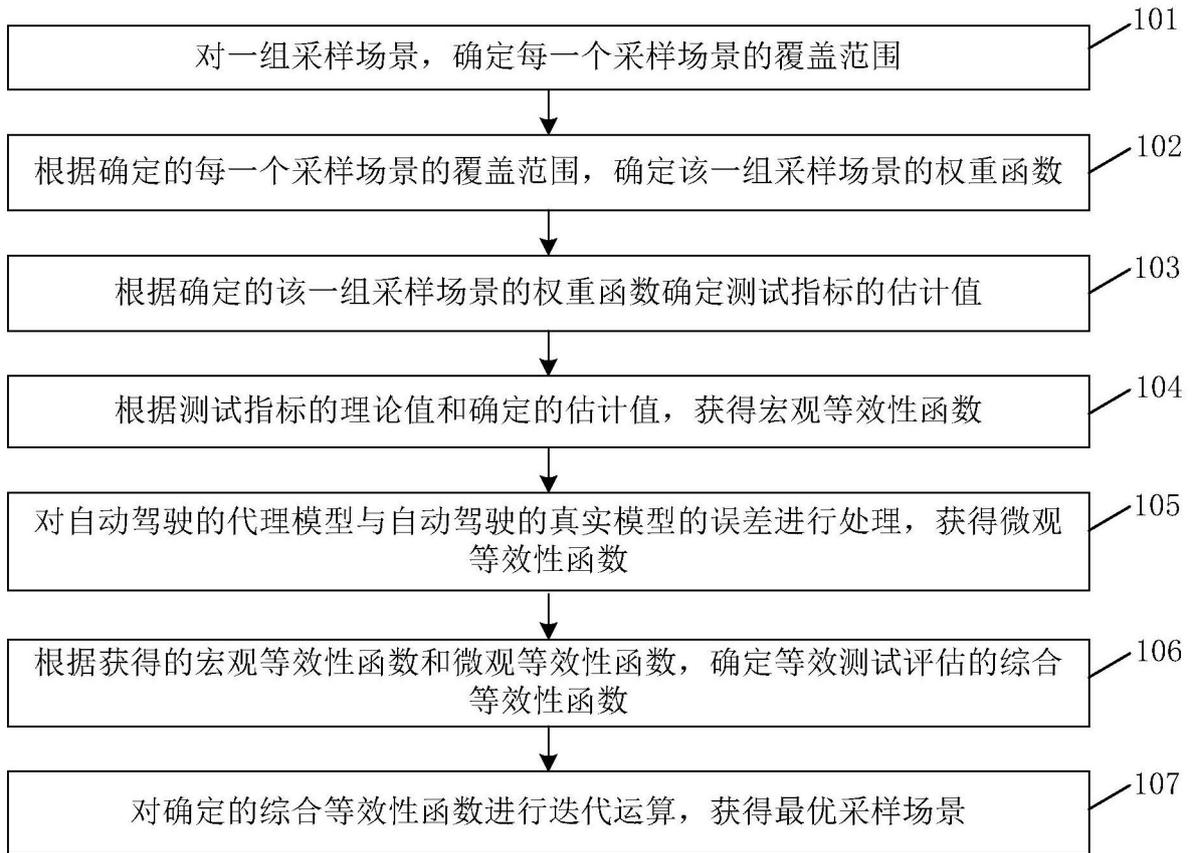


图1

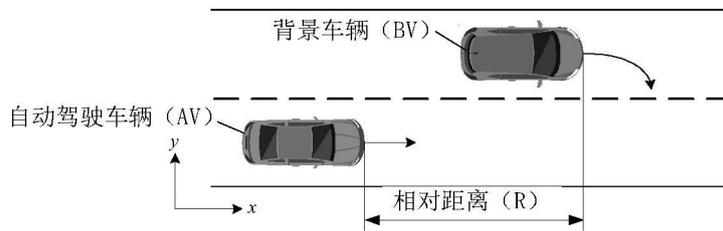


图2

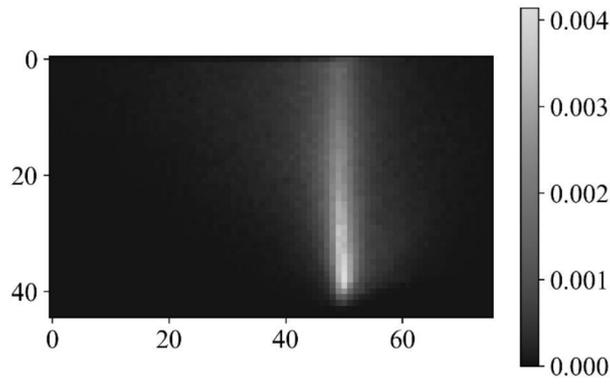


图3

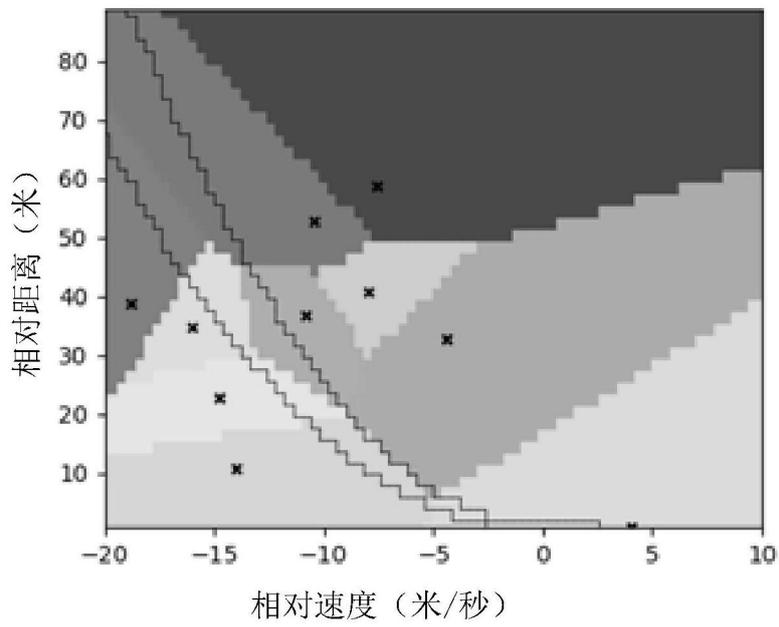


图4

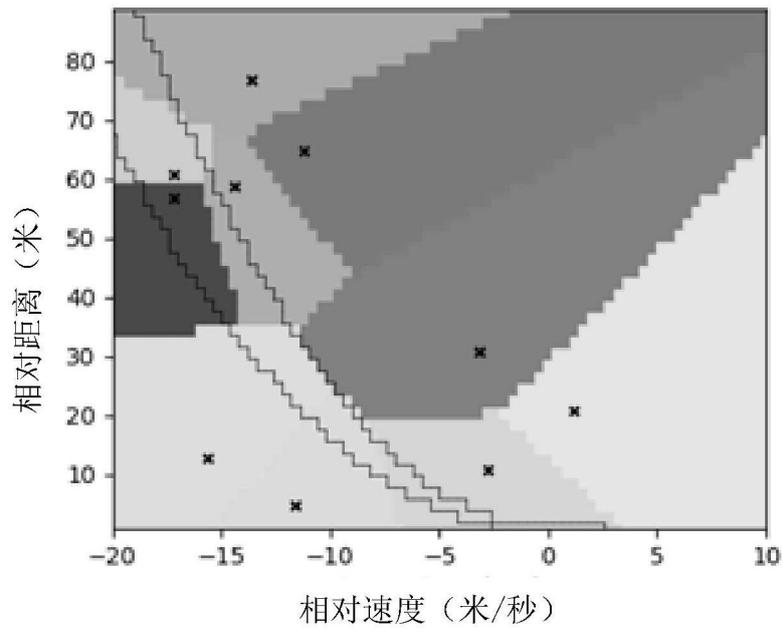


图5