



**República Federativa do Brasil**

Ministério do Desenvolvimento, Indústria,  
Comércio e Serviços

Instituto Nacional da Propriedade Industrial



**(11) BR 102016005478-8 B1**

**(22) Data do Depósito:** 11/03/2016

**(45) Data de Concessão:** 30/05/2023

**(54) Título:** MÉTODO IMPLEMENTADO POR COMPUTADOR PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS E SISTEMA PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS

**(51) Int.Cl.:** G06N 7/00; G05B 13/02; G05B 13/04; G06N 5/04.

**(52) CPC:** G06N 7/005; G05B 13/027; G05B 13/04; G06N 5/047.

**(30) Prioridade Unionista:** 15/10/2015 IN 3922/MUM/2015.

**(73) Titular(es):** TATA CONSULTANCY SERVICES LIMITED.

**(72) Inventor(es):** KARAMJIT SINGH; GAUTAM SHROFF; PUNEET AGARWAL.

**(57) Resumo:** MÉTODO IMPLEMENTADO POR COMPUTADOR PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS, SISTEMA PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS, E, PRODUTO DE PROGRAMA DE COMPUTADOR A presente invenção se refere a sistemas e a métodos para a mineração de fiabilidade previsível que permitem a previsão de falhas emergentes inesperadas no futuro sem esperar pelo início da ocorrência de falhas reais em números significativos. São identificados conjuntos de Códigos de Diagnóstico de Falhas (DTCs) discriminativos de máquinas conectadas em uma população antes da falha das partes associadas. É gerado um modelo de dependência condicional temporal com base na dependência temporal entre a falha das partes a partir de dados relativos a falhas passadas e de conjuntos de DTCs discriminativos identificados. As falhas futuras são previstas com base em uma dependência condicional temporal gerada e é realizada uma análise das causas subjacentes às falhas futuras previstas para a mineração de fiabilidade previsível. A probabilidade de falha é computadorizada com base em ambas, na ocorrência e na não-ocorrência de DTCs. A análise das causas subjacentes permite identificar um subconjunto da população quando é gerado um aviso precoce e também quando não é (...).

MÉTODO IMPLEMENTADO POR COMPUTADOR PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS E SISTEMA PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS

REFERÊNCIA CRUZADA A PEDIDOS RELACIONADOS E PRIORIDADE

[001] O presente pedido reivindica a prioridade do Pedido de Patente Indiano n.º 3922/MUM/2015 apresentado em 15 de outubro, 2015, a totalidade do qual é incorporada por referência pelo presente.

CAMPO TÉCNICO

[002] As formas de realização no presente documento se referem à mineração de fiabilidade previsível e, mais particularmente, a sistemas e a métodos envolvendo modelos de fiabilidade aumentada por sensores.

ANTECEDENTES

[003] Com a internet industrial a desenhar o futuro é apenas natural ter máquinas conectadas formando parte de qualquer aspecto de tecnologia. Tradicionalmente, a mineração de fiabilidade previsível tem vindo a ser baseada em dados históricos relativos a falhas parciais de reivindicações de garantia usando distribuições de família exponencial tais como a distribuição de Weibull ou de log-normal. Quando as falhas observadas (em uma ou mais partes) em uma população de máquinas excede o número esperado com base em um modelo desta natureza, isto pode servir como um aviso precoce de um eventual problema sistêmico na população. Estes avisos precoces assentam em algumas falhas excepcionalmente elevadas tendo ocorrido realmente. Mais uma vez, tem vindo a ser considerado que desvios significativos

de contagens de falhas esperadas frequentemente apenas ocorrem em algum subconjunto desconhecido da população, por exemplo, um lote específico, ou máquinas fabricadas em um ano específico ou em uma instalação específica e similar. Estes desvios são insignificantes na população total e permanecem não identificados quando são usadas as técnicas de mineração de fiabilidade tradicional. É um desafio não detectar apenas eventuais problemas mais cedo do que uma eventual análise usando a fiabilidade tradicional mas igualmente identificar um subconjunto da população em que possa ter ocorrido uma anomalia que de outra forma estaria escondida estatisticamente na população.

#### SUMÁRIO

[004] Este sumário é providenciado para introduzir conceitos relativos à mineração de fiabilidade previsível. No contexto da presente invenção, a expressão "máquinas conectadas" se refere a veículos conectados, motores conectados ou qualquer outro equipamento industrial que serve como máquina conectada conhecida no estado da técnica. Os sistemas de diagnóstico de bordo (OBD) providenciados para monitorar partes das máquinas conectadas incluem sensores no contexto da presente invenção. Este sumário não pretende identificar características fundamentais do objeto reivindicado nem pretende ser usado para determinar ou limitar o âmbito da presente invenção.

[005] Os sistemas e os métodos da presente invenção permitem a mineração de fiabilidade previsível em uma população de máquinas conectadas. Devido à "internet industrial", a maior parte do equipamento industrial, instala sensores que monitoram continuamente o comportamento do tempo

de execução de componentes desejados em campo e transmitem ainda de volta uma informação de sensor predefinida ao fabricante por vários meios incluindo redes celulares sem fios ou de WiFi metropolitanas. Em particular, os veículos automóveis modernos apresentam módulos de controle eletrônico de bordo que geram Códigos de Diagnóstico de Falhas (DTCs) alfanuméricos para indicar níveis de sensor anormais em várias situações, algumas das quais são indicativas de uma falha de parte real ou eventual. Estes DTCs tipicamente são desencadeados antes da falha de parte real. Os sistemas e os métodos da presente invenção analisam estes DTCs para correlacionar os DTCs como momentos de falha futuros para servirem como indicadores de aviso precoce de eventuais falhas de partes futuras.

[006] Em um aspecto é providenciado um método implementado por computador para a mineração de fiabilidade previsível em uma população de máquinas conectadas, em que o método compreende identificar conjuntos de Códigos de Diagnóstico de Falhas (DTCs) discriminativos a partir de DTCs gerados por falhas passadas, em que os conjuntos de DTCs discriminativos correspondem a partes predefinidas associadas das máquinas conectadas; gerar um modelo de dependência condicional temporal com base na dependência temporal entre a falha das partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados, e; prever falhas futuras com base em uma dependência condicional temporal gerada e na ocorrência e na não-ocorrência de DTCs.

[007] Em uma forma de realização, a etapa de prever falhas futuras pode ser seguida por realização de uma

análise das causas subjacentes às falhas futuras previstas para a mineração de fiabilidade previsível.

[008] Em uma forma de realização, a etapa de identificar conjuntos de DTCs discriminativos é baseada na mineração de regras de associação, em que a mineração de regras de associação compreende o uso da técnica Apriori.

[009] Em uma forma de realização, os antecedentes das regras identificadas pela técnica de mineração de regras de associação formam o conjunto de DTCs discriminativos.

[010] Em uma forma de realização, o modelo de dependência condicional temporal é uma rede Bayesiana.

[011] Em uma forma de realização, a dependência temporal entre a falha das partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados segue uma distribuição de Weibull.

[012] Em uma forma de realização, a etapa de gerar um modelo de dependência condicional temporal é seguida por uma etapa de separar a população de máquinas conectadas em um primeiro conjunto compreendendo máquinas conectadas em que os DTCs não são gerados em um período de tempo predeterminado e um segundo conjunto compreendendo máquinas conectadas em que pelo menos um dos DTCs é gerado no período de tempo predeterminado.

[013] Em uma forma de realização, a etapa de prever falhas futuras compreende a etapa de processar por computador a probabilidade de falha com base em ambas, na ocorrência e na não-ocorrência de DTCs, na população separada de máquinas conectadas e gerar um aviso precoce quando o

número previsto de falhas é superior ao número esperado de falhas com base nos dados relativos a falhas passadas por um valor predefinido. Em uma forma de realização, o valor predefinido é baseado no número previsto de falhas e a variância de uma variável aleatória representando o número de falhas das partes predefinidas em um período de tempo predeterminado.

[014] Em uma forma de realização, a etapa de realizar uma análise das causas subjacentes compreende identificar um subconjunto da população quando é gerado um aviso precoce para pelo menos uma das partes predefinidas. Em uma forma de realização, a etapa de realizar uma análise das causas subjacentes compreende: (i) calcular um primeiro momento de falha esperado com base nos dados relativos às falhas passadas; (ii) separar a população de máquinas conectadas em um primeiro conjunto compreendendo máquinas conectadas em que os DTCs não são gerados em um período de tempo predeterminado e um segundo conjunto compreendendo máquinas conectadas em que pelo menos um DTC é gerado no período de tempo predeterminado; (iii) calcular um segundo momento de falha esperado com base no momento de falha esperado do segundo conjunto que é baseado ainda em parâmetros retardadores predefinidos e o momento de ocorrência de pelo menos um DTC; (iv) definir uma pontuação de anomalia para cada uma das máquinas conectadas com base no primeiro momento de falha esperado e no segundo momento de falha esperado calculados; (v) realizar iterativamente as etapas desde (i) até (iv) para características predefinidas das máquinas conectadas, e; (vi) identificar o subconjunto da população apresentando a pontuação de anomalia superior a um

limiar predefinido, em que o subconjunto identificado indica eventuais razões para o aviso precoce para cada uma das características predefinidas.

[015] Em uma forma de realização, a etapa de realizar uma análise das causas subjacentes compreende identificar um subconjunto da população com uma eventual anomalia quando não é gerado um aviso precoce a nível da população. Em uma forma de realização, a etapa de realizar uma análise das causas subjacentes compreende: (i) definir uma pontuação de anomalia para cada uma das máquinas conectadas; (ii) associar cada uma das máquinas conectadas a um registro compreendendo um conjunto de características predefinidas e a pontuação de anomalia definida; (iii) discretizar a pontuação de anomalia definida em um nível elevado predefinido ou em um nível normal; (iv) realizar uma mineração de regras de associação para identificar regras com a pontuação de anomalia de nível elevado; (v) agregar as regras de associação identificadas usando uma técnica com base na densidade para formar agregados de regras; (vi) selecionar uma ou mais regras de cada um dos agregados de regras que apresentam suporte e confiança elevados; e (vii) identificar o subconjunto da população com a pontuação de anomalia de nível elevado com base em antecedentes das uma ou mais regras selecionadas que são indicativas de eventuais razões para a pontuação de anomalia de nível elevado.

[016] Em outro aspecto é providenciado um sistema para a mineração de fiabilidade previsível em uma população de máquinas conectadas, em que o sistema compreende: um ou mais processadores; um dispositivo de interface de comunicação; um ou mais dispositivos de

armazenamento de dados internos acoplados operacionalmente com um ou mais processadores para o armazenamento: um módulo de entrada configurado para receber Códigos de Diagnóstico de Falhas (DTCs) a partir de sistemas de diagnóstico de bordo de partes predefinidas das máquinas conectadas; um identificador de padrões de DTC configurado para identificar conjuntos de DTCs discriminativos a partir dos DTCs gerados por falhas precedentes, em que os conjuntos de DTCs discriminativos correspondem a partes predefinidas associadas das máquinas conectadas; um gerador de redes Bayesianas configurado para gerar um modelo de dependência condicional temporal com base na dependência temporal entre a falha de partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados; e um dispositivo de previsão de falhas configurado para prever falhas futuras com base na dependência condicional temporal gerada e na ocorrência e na não-ocorrência de DTCs.

[017] Em uma forma de realização, o sistema acima descrito pode compreender ainda um analisador configurado para realizar uma análise das causas subjacentes às falhas futuras previstas para a mineração de fiabilidade previsível.

[018] Em ainda outro aspecto é providenciado um produto de programa de computador para processar dados, compreendendo um meio legível por computador não-transitório apresentando instruções incorporadas neste para identificar conjuntos de Códigos de Diagnóstico de Falhas (DTCs) discriminativos a partir de DTCs gerados por falhas precedentes, em que os conjuntos de DTCs discriminativos correspondem a partes predefinidas associadas das máquinas



conectadas; gerar um modelo de dependência condicional temporal com base na dependência temporal entre a falha das partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados; prever falhas futuras com base na dependência condicional temporal gerada e na ocorrência e na não-ocorrência de DTCs; e realizar uma análise das causas subjacentes das falhas futuras previstas para a mineração de fiabilidade previsível.

#### BREVE DESCRIÇÃO DAS FIGURAS

[019] As formas de realização no presente documento serão mais bem compreendidas com base na descrição detalhada com referência às figuras, em que a:

[020] FIG. 1 ilustra um diagrama de blocos exemplificativo de um sistema para a mineração de fiabilidade previsível em uma população de máquinas conectadas de acordo com uma forma de realização da presente invenção;

[021] FIG. 2 é um fluxograma exemplificativo ilustrando um método implementado por computador para a mineração de fiabilidade previsível em uma população de máquinas conectadas usando o sistema da FIG. 1 de acordo com uma forma de realização da presente invenção;

[022] FIG. 3 é um modelo de fiabilidade Bayesiana aumentado por sensores de acordo com uma forma de realização da presente invenção;

[023] FIG. 4 ilustra gráficos (quantil-quantil) QQ de DTCs retirados de uma distribuição de Weibull e de Gaussian de um conjunto de dados exemplificativo;

[024] FIG. 5 ilustra uma comparação do número de falhas previstas por SARM versus real versus

esperado por BRM;

[025] FIG. 6 ilustra a tendência semanal de falhas e de previsões para a parte P<sub>G</sub>;

[026] FIG. 7 ilustra uma análise das causas subjacentes ao aviso precoce (EW-1) usando ferramentas de análise visual, e;

[027] FIG. 8 ilustra um subgrupo descoberto para aviso precoce (EW-2) como um par de exceção à regra.

[028] Para o perito na técnica é evidente que qualquer diagrama de blocos no presente documento representa vistas conceptuais de sistemas ilustrativos de realização dos princípios do objeto da presente invenção. Da mesma forma, é evidente que quaisquer gráficos, fluxogramas, diagramas de transição de estado, pseudocódigos e similares representam vários processos que essencialmente podem ser representados em um meio legível por computador e, por conseguinte, executados por um dispositivo de computação ou processador, quer ou não este dispositivo de computação ou processador seja apresentado explicitamente.

#### DESCRIÇÃO DETALHADA

[029] As formas de realização no presente documento e as várias características e detalhes vantajosos das mesmas são mais detalhadamente explicados com referência às formas de realização não limitativas que são ilustradas nas figuras em anexo e detalhados na descrição que se segue. Os exemplos usados no presente documento apenas pretendem facilitar a compreensão das formas de colocar em prática as formas de realização no presente documento e ainda para habilitar o perito na técnica a colocar em prática as formas de realização no presente documento. Assim, os exemplos não

devem ser considerados como limitativos do âmbito das formas de realização no presente documento.

[030] As palavras "compreendendo", "apresentando", "contendo" e "incluindo" e outras formas correspondentes apresentam um significado equivalente e são abertas no sentido em que o item ou os itens que se seguem a qualquer uma destas palavras não pretendem ser uma listagem exaustiva desse item ou itens ou estar limitados apenas ao item ou itens alistados.

[031] É igualmente evidente que conforme usadas no presente documento e nas reivindicações anexas, as formas singulares "um", "um" e "o" incluem as referências plurais exceto se expresso claramente de outra forma pelo contexto. Embora possam ser usados quaisquer sistemas e métodos similares ou equivalentes àqueles descritos no presente documento na prática ou no teste das formas de realização da presente invenção, os sistemas e os métodos preferidos são descritos em seguida.

[032] Algumas formas de realização da invenção, ilustrando todas as suas características, serão mais detalhadamente discutidas em seguida. As formas de realização divulgadas são apenas exemplificativas da invenção, que pode ser realizada de várias formas.

[033] Antes de avançar com a explicação detalhada é de salientar que toda a discussão abaixo, independentemente da implementação específica que está a ser descrita, é de natureza exemplificativa e não limitativa. Por exemplo, embora aspectos, características ou componentes selecionados das implementações sejam apresentados como sendo armazenados em memórias, a totalidade ou uma parte dos

sistemas e dos métodos em conformidade com o sistema e com o método de reconhecimento e de classificação de equipamento podem ser armazenados, distribuídos ou lidos a partir de quaisquer outros meios legíveis por máquinas.

[034] No que se refere às figuras e, mais particularmente às FIGS. de 1 a 8, em que os mesmos números de referência se referem a características correspondentes em todas as figuras, são apresentadas formas de realização preferidas e estas formas de realização são descritas no contexto do sistema e do método exemplificativo que se segue.

[035] A FIG. 1 ilustra um diagrama de blocos exemplificativo de um sistema 100 para a mineração de fiabilidade previsível de acordo com uma forma de realização da presente invenção e a FIG. 2 ilustra um fluxograma exemplificativo ilustrando um método implementado por computador 200 para o modelo de fiabilidade previsível usando o sistema da FIG. 1 de acordo com uma forma de realização da presente invenção. As etapas do método 200 da presente invenção serão agora explicadas com referências aos componentes do sistema 100 conforme apresentado na FIG. 1. Em uma forma de realização, o sistema 100 inclui um ou mais processadores (não apresentados), uma interface de comunicação ou interface de entrada/saída (I/O) (não apresentadas) e uma memória ou um ou mais dispositivos de armazenamento de dados internos (não apresentados) acoplados operacionalmente com o um ou mais processadores. O um ou mais processadores podem ser implementados como um ou mais microprocessadores, microcomputadores, microcontroladores, processadores de sinais digitais, unidades de processamento central, máquinas de estado, circuitos lógicos e/ou quaisquer

dispositivos que manipulem sinais com base em instruções operacionais. Entre outras capacidades, o processador é configurado para aceder a e executar instruções legíveis por computador armazenadas na memória. Em uma forma de realização, o sistema 100 pode ser implementado em um servidor ou em uma variedade de sistemas de computação, tais como um computador portátil, um computador fixo, um *notebook*, uma estação de trabalho, um computador do tipo *mainframe*, um servidor, um servidor de rede, uma nuvem, um dispositivo portátil e similar.

[036] A interface I/O pode incluir uma pluralidade de interfaces de software e de hardware, por exemplo, uma interface web, uma interface de usuário gráfica e similares e pode facilitar comunicações múltiplas em uma vasta gama de tipos de rede e de protocolo, incluindo redes por cabo, por exemplo, LAN, cabo, etc., e redes sem fios, tais como WLAN, celular ou satélite. Em uma forma de realização, a interface I/O pode incluir uma ou mais portas para conectar um número de dispositivos uns com os outros ou com outro servidor.

[037] A memória pode incluir qualquer meio legível por computador conhecido na técnica incluindo, por exemplo, memória volátil, tal como memória estática de acesso aleatório (SRAM) e memória dinâmica de acesso aleatório (DRAM) e/ou memória não volátil, tal como memória só de leitura (ROM), ROM programável delével, memórias *flash*, discos rígidos, discos ópticos e fitas magnéticas. Em uma forma de realização, os vários módulos do sistema 100 podem ser armazenados na memória.

[038] As máquinas conectadas são providenciadas

com sistemas de diagnóstico de bordo (OBD) 110. Tipicamente existe um mapeamento exagerado entre os Códigos de Diagnóstico de Falhas (DTCs) e partes das máquinas conectadas, quer dizer, um DTC pode indicar uma falha em mais do que uma parte e uma falha em uma parte pode desencadear mais do que um DTC. Por conseguinte, é importante identificar DTCs discriminativos que possam ser correlacionados com falhas futuras de cada parte, mesmo quando não estão associadas apenas a falhas de uma parte única. Na etapa 202, um identificador de padrões DTC 114 pode identificar conjuntos de DTCs discriminativos em que os conjuntos de DTCs discriminativos correspondem a partes predefinidas associadas das máquinas conectadas. Em uma forma de realização, a etapa de identificar conjuntos de DTCs discriminativos é baseada na mineração de regras de associação. Em uma forma de realização, a mineração de regras de associação pode incluir identificar regras para DTCs discriminativos para cada parte por análise das ocorrências de DTC em situações que levam a falhas e por comparação destas com situações em que não são observadas falhas.

[039] Para usar os DTCs eficazmente como um indicador de aviso precoce de uma falha de parte, na etapa 204, é gerado um modelo de dependência condicional temporal por um gerador de redes Bayesianas 116. Em uma forma de realização, o modelo de dependência condicional temporal é baseado na dependência temporal entre a falha das partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados. Os sistemas e os métodos da presente invenção usam os dados relativos a falhas passadas, os conjuntos de DTCs

discriminativos identificados e a dependência temporal para aperfeiçoar os modelos de fiabilidade convencionais. Em uma forma de realização, a dependência temporal entre a falha de partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados segue uma distribuição de Weibull.

[040] Na etapa 206, as falhas futuras são previstas com base na dependência condicional temporal gerada por um dispositivo de previsão de falhas 118. Em uma forma de realização, a etapa de prever falhas futuras compreende gerar um aviso precoce quando o número previsto de falhas é superior ao número esperado de falhas com base nos dados relativos a falhas passadas por um valor predefinido. Em uma forma de realização, o dispositivo de previsão de falhas 118 é configurado para identificar partes das máquinas conectadas associadas ao aviso precoce e correspondentemente identificar um subconjunto da população associado.

[041] Na etapa 208 é realizada uma análise das causas subjacentes às falhas futuras previstas para a mineração de fiabilidade previsível por um analisador 120.

[042] As etapas de 202 a 208 da FIG. 2 serão mais detalhadamente explicadas em seguida. No contexto da presente invenção, as várias expressões usadas doravante são definidas conforme apresentadas abaixo.  $I = \{i_1, i_2, \dots, i_m\}$  é um conjunto de atributos de binário  $m$  designados por itens.  $D = \{t_1, t_2, \dots, t_n\}$  é um conjunto de transações designado por base de dados. Cada transação em  $D$  apresenta uma id única e contém um subconjunto de itens em  $I$ . Os conjuntos de itens  $X$ ,  $Y$  são designados por antecedente e por consequente de uma regra, respectivamente. Uma regra de associação  $r$  é uma

expressão de implicação da forma  $X \rightarrow Y$ , em que  $X, Y \subseteq I$  e  $X \cap Y = \emptyset$ . O suporte  $s(X)$  de um conjunto de itens  $X$  é definido como a fracção de transações na base de dados  $D$  que contém o conjunto de itens  $X$ . A confiança de uma regra  $conf(r)$  é  $s(X \cup Y) / s(X)$ . Uma elevação adicional de uma regra  $l(r)$  mede o respectivo interesse e é uma relação da respectiva confiança e suporte do conseqüente. Quer dizer,  $conf(r)/s(Y)$ . A cobertura de uma regra  $r$  é a fracção do número de transação em  $D$  contendo  $X$  e dada por  $s(X)$ .  $R = \{r_1, r_2, \dots, r_l\}$  é um conjunto de regras  $l$ , então a cobertura de  $R$  representada por  $c_R$  em  $D$  é uma fracção de transações em  $D$  contendo os antecedentes de regras no conjunto  $R$ .

[043] A etapa 202 em que são identificados conjuntos de Códigos de Diagnóstico de Falhas (DTCs) discriminativos envolve identificar DTCs que são discriminativos para falhas da parte  $P_i$ , quer dizer, os DTCs que ocorrem antes da falha de  $P_i$  mas não antes de outras falhas de partes. Em uma forma de realização, identificar DTCs discriminativos é baseado na mineração de regras de associação. Em uma forma de realização, a mineração de regras de associação usa a técnica Apriori para identificar regras de DTCs que conduzem à falha da parte  $P_i$  com confiança e elevação elevadas. Considerando que uma parte única pode falhar devido a diferentes razões em diferentes veículos é possível obter mais do que uma regra de confiança elevada para uma parte única. Para a parte  $P_i$ , em que  $I$  é o conjunto de itens de todos os DTCs possíveis para a totalidade da população de máquinas conectadas. Em qualquer momento  $t_0$ ,  $P_i$  representa uma parte falhada em  $n$  máquinas conectadas. Para cada uma das  $n$  máquinas conectadas, partindo da data de falha



da parte  $P_i$ , todos os DTCs em transação  $f_i$  que são desencadeados nos dias  $d$  passados são recolhidos e o item  $f$  representando a falha é adicionado a cada transação. Por conseguinte, para  $n$  máquinas conectadas, é obtido um conjunto de  $n$  transações  $D_f = \{f_1, f_2, \dots, f_n\}$  designado por conjunto de falhas, em que cada  $f_i \subseteq I$ . Da mesma forma, é obtido um outro conjunto  $D_{nf}$  designado por conjunto de não falhas para as máquinas conectadas, em que partes diferentes de  $P_i$  apresentaram falhas e um item  $n_f$  é adicionado a cada transação do conjunto  $D_{nf}$ . Para a ocorrência múltipla de um DTC único em uma transação é considerada a primeira ocorrência desse DTC. É gerado um conjunto de regras  $R = D_f \cup D_{nf}$  usando a técnica Apriori. São identificadas as regras no conjunto  $R$  apresentando uma confiança com  $f(r) > \tau_p \forall r \in R$ , e  $c_R$  representando a cobertura do conjunto  $R$  em  $D_f > R_c$ . Neste caso,  $\tau_p$  é um limiar preliminar para a confiança e  $R_c$  é um limiar preliminar para a cobertura. O conjunto de antecedentes de todas as regras no conjunto  $R$  formam o conjunto de DTCs discriminativos conforme  $disc_i = \{D_1, D_2, D_3, \dots, D_N\}$  para a parte  $P_i$ , em que cada conjunto  $D_j$  contém pelo menos um DTC e  $D_1 \cap D_2 \cap \dots \cap D_N = \emptyset$ .

[044] Considerando que os DTCs ocorrem antes da falha real, o retardamento temporal entre a ocorrência do DTC e a falha da parte define a dependência entre estes. Na etapa 204 é gerado um modelo de fiabilidade Bayesiana ou uma rede Bayesiana com base na dependência entre a falha da parte  $P_i$  e o respectivo conjunto de DTCs discriminativos  $disc_i$  conforme ilustrado na FIG. 3. Cada nó da rede Bayesiana representa uma variável aleatória contínua conforme descrita abaixo. No contexto da presente invenção, a variável  $t_{pi}$  se refere ao

momento em que a parte  $P_i$  falha. Em uma forma de realização, o momento  $t_{pi}$  segue a distribuição de Weibull com parâmetros de falha  $\alpha_i$  e  $\beta_i$ , representando parâmetros de escala e de forma da distribuição de Weibull, respectivamente conforme representados na equação (1) abaixo.

$$P(t_{pi}|\alpha_i, \beta_i) \sim Weibull(\alpha_i, \beta_i) \quad (1)$$

[045] No contexto da presente invenção, a distribuição uniforme é assumida como um pré-requisito para  $\alpha_i$  e  $\beta_i$  com um limite inferior de 0 e limite superior de  $a > 0$  e  $b > 0$  respectivamente, quer dizer  $P(\alpha_i) \sim U(0, a)$  e  $P(\beta_i) \sim U(0, b)$ .

[046] O nó  $\tau_i$  na FIG. 3 representa a variável  $\tau_{ij}$  que se refere ao retardamento temporal entre o momento de falha da parte  $P_i$  e o momento de ocorrência de DTC  $D_j$ . Em uma forma de realização, o momento  $\tau_{ij}$  segue a distribuição de Weibull com os parâmetros de retardamento  $\alpha_{Dj}$  e  $\beta_{Dj}$ , representando parâmetros de escala e de forma da distribuição de Weibull, respectivamente. Os parâmetros de retardamento capturam a dependência condicional temporal entre a parte  $P_i$  e DTC  $D_j$  conforme representada na equação (2) abaixo.

$$P(\tau_{ij}|\alpha_{Dj}, \beta_{Dj}) \sim Weibull(\alpha_{Dj}, \beta_{Dj}) \quad (2)$$

[047] No contexto da presente invenção, a distribuição uniforme é assumida como um pré-requisito para os parâmetros de retardamento com um limite inferior 0 e um limite superior de  $a_{Dj} > 0$  e  $b_{Dj} > 0$  respectivamente, quer dizer  $P(\alpha_{Dj}) \sim U(0, a_{Dj})$  e  $P(\beta_{Dj}) \sim U(0, b_{Dj})$ .

[048] Considerando que os DTCs ocorrem antes da falha da parte, no contexto da presente invenção, o nó  $t_D$  representa a variável  $t_{Dj}$  que se refere ao momento em que ocorre DTC  $D_j$  em termos de variáveis  $t_{pi}$  e  $\tau_{ij}$ . Em uma forma

de realização, o momento  $t_{D_j}$  segue uma distribuição normal com uma média  $t_{p_i} - \tau_{ij}$  com um desvio padrão  $\sigma_j$  conforme representado na equação (3) abaixo.

$$P(t_{D_j}|t_{p_i}, \tau_{ij}, \sigma_j) \sim N(t_{p_i} - \tau_{ij}, \sigma_j) \quad (3)$$

em que um pré-requisito para  $\sigma_j$  é a distribuição uniforme, quer dizer  $\sigma_j \sim U(0, s_j)$ .

[049] De acordo com a presente invenção, a rede Bayesiana se aproxima da distribuição com um nó representado por  $t_{p_i}$ , um nó para cada DTC  $D_j$  no conjunto discriminativo de  $P_i$  representado por  $\tau_{ij}$ , um para cada  $D_j \in \text{disc}_i$  representado por  $t_{D_j}$  e o resto dos nós representando os parâmetros destas três variáveis aleatórias. Todos os parâmetros do modelo, nomeadamente os parâmetros de Weibull para as falhas e os DTCs e a média e a variância dos retardamentos distribuídos normalmente são determinados usando a amostragem do tipo Markov Chain Monte Carlo (MCMC) relativa aos dados relativos a falhas passadas e aos dados relativos ao DTC de uma população de máquinas conectadas.

[050] Na etapa 206, as falhas futuras são previstas com base nos dados relativos a falhas passadas e nos DTCs discriminativos que foram modulados como rede Bayesiana na etapa 204.  $N_{p_i}$  representa o número de falhas da parte  $P_i$  durante um intervalo de tempo específico no futuro com base em DTCs e  $M_{p_i}$  representa o número de falhas esperado usando o modelo de fiabilidade convencional. Comparando estatisticamente  $N_{p_i}$  e  $M_{p_i}$  é gerado um aviso precoce pelo sistema 100 quando  $N_{p_i}$  é superior a  $M_{p_i}$  por um valor predeterminado.

[051] Considerando  $m$  máquinas conectadas e o momento  $t_0$ , supondo que  $P_i$  falhou em  $n$  de  $m$  máquinas

conectadas,  $V$  é o conjunto de máquinas conectadas restantes  $r$  ( $= m - n$ ), a probabilidade de falha de a parte  $P_i$  falhar em  $[t_1, t_2]$  ( $t_2 > t_1 > t_0$ ) considerando que sobreviveu até  $t_0$  é representada pela equação (4) abaixo.

$$P(t_{pi} \in [t_1, t_2] | t_{pi} > t_0) = \frac{P(t_{pi} \leq d_2) - P(t_{pi} \leq d_1)}{P(t_{pi} > t_0)} = \frac{F(d_2) - F(d_1)}{S(t_0)} \quad (4)$$

em que  $d_k$  é o momento calculado até  $t_k$  a partir de um momento inicial e  $F(t)$  é a probabilidade de que a parte vai falhar antes do momento  $t$ .

[052]  $F(t)$ , conforme conhecido na técnica, pode ser representado pela equação (5) abaixo.

$$F(t) = 1 - e^{-(t/\alpha)^\beta} \quad (5)$$

em que  $\alpha$  e  $\beta$  são parâmetros de escala e de forma da distribuição de Weibull e  $S(t)$  é a chamada probabilidade de sobrevivência que pode ser representada pela equação (6) abaixo.

$$S(t) = 1 - F(t) \quad (6)$$

[053] A probabilidade de falha da parte  $P_i$  para cada  $v \in V$  é calculada usando o modelo de fiabilidade básica (BRM) convencional assim como o modelo de fiabilidade aumentada por sensores (SARM) que incorpora DTCs. No BRM convencional apenas são usados dados relativos a falhas passadas da parte  $P_i$  para calcular a probabilidade de falhas em  $r$  veículos. Usando os "parâmetros de falhas" enquanto parâmetros de escala e de forma e considerando o momento  $t_s^v$  de cada veículo  $v \in V$  como momento inicial, quer dizer,  $d_k^v = t_k - t_s^v$  nas equações 4 e 5. Então é calculado  $P_i^v$ , a

probabilidade de falha para cada  $v \in V$ .

[054] No SARM da presente invenção são usados os dados relativos a falhas passadas e os DTCs. Os veículos  $r$  restantes em que a parte  $P_i$  não falhou até  $t_0$  são divididos em dois conjuntos 1)  $V_1$ : conjunto de veículos em que ocorreu pelo menos um de  $D_j \in \text{disc}_i$  em  $[t_0 - d, t_0]$ , e, 2)  $V_0$ : conjunto de veículos em que não ocorreu nenhum de  $D_j \in \text{disc}_i$  em  $[t_0 - d, t_0]$ .

[055] Para os veículos em  $V_1$ , considerando que ocorreu pelo menos um de  $D_j \in \text{disc}_i$  em  $[t_0 - d, t_0]$ , para cada  $v \in V_1$ , o momento de ocorrência  $t_{D_j}$  do DTC  $D_j$  é usado como o momento inicial, quer dizer  $d_k = t_k - t_{D_j}$ . Usando ainda os "parâmetros de retardamento" como parâmetros de escala e de forma nas equações 4 e 5,  $P_{i1}^v$ , a probabilidade de falha da parte  $P_i$  para cada  $v \in V_1$  é calculada considerando que DTC  $D_j$  ocorreu no momento  $t_{D_j}$ . No caso em que tenha ocorrido mais do que um DTC discriminativo em um veículo é usado um DTC de confiança elevada. Além disso, considerando que a confiança  $\text{conf}(r)$  da regra  $r: D_j \rightarrow P_i$  afirma que  $(1 - \text{conf}(r))\%$  vezes DTC  $D_j$  levará à falha diferente de  $P_i$ ,  $P_{i1}^v$  é marginalizado com a equação 7 abaixo.

$$P_{i1}^v = \text{conf}(r)P_{i1}^v + (1 - \text{conf}(r))P_i^v \quad (7)$$

em que,  $P_i^v$  é a probabilidade calculada para cada  $v \in V_1$  tal como no caso do BRM.

[056] Para os veículos em  $V_0$  é calculada a probabilidade de falha para a parte  $P_i$  para cada  $v \in V_0$  usando apenas os dados relativos a falhas passadas, quer dizer usando os "parâmetros de falha" como parâmetros de

escala e de forma nas equações 4 e 5 é calculado o  $p_0^v$  para cada  $v \in V_0$ .  $t_s^v$ , o momento de venda de um veículo  $v \in V_0$  é usado como momento inicial, quer dizer  $d_k^v = t_k - t_s^v$ . Considerando que para cada  $v \in V_0$  não ocorreu nenhum DTC  $D_j \in \text{disc}_i$  em  $[t_0 - d, t_0]$ , mas como a definição de  $c_R$  afirma que  $c_R\%$  de vezes, ocorrerá pelo menos um de  $D_j \in \text{disc}_i$  em  $[t_{pi} - d, t_{pi}]$  antes da falha da parte  $P_i$ , em que a probabilidade  $p_0^v$  usando  $c_R$  é marginalizada com a equação 8 abaixo.

$$p_{i0}^v = p_{i0}^v(1 - c_R)(1 - p_{ij}) + p_{i0}^v c_R p_{ij} \quad (8)$$

em que  $p_{ij}$  é a probabilidade de que ocorrerá pelo menos um de DTC  $D_j \in \text{disc}_i$  em  $[t_0, t_1]$ . Subsequentemente é determinado o  $p_{ij}$  para a parte  $P_i$  usando a falha  $n$  da parte  $P_i$  que ocorreu até  $t_0$ .

[057] Para determinar se o volume de falhas conforme previsto pelo modelo BRM e pelo modelo SARM da presente invenção diferem o suficiente para declarar um aviso precoce, a falha da parte  $P_i$  em um veículo é considerada como uma variável aleatória da distribuição de Bernoulli com o parâmetro  $p$  (probabilidade de falha de  $P_i$ ). Considerando que a probabilidade de falha de  $P_i$  é diferente de veículo para veículo, as falhas de  $P_i$  em  $r$  veículos formam as variáveis aleatórias da distribuição de Bernoulli independentes de  $r$  e não idênticas. A soma destas variáveis  $r$  forma uma outra variável aleatória  $X$  representando o número de falhas de  $P_i$  em  $[t_1, t_2]$  que segue a distribuição binomial de Poisson. A media de  $X$  para ambos os casos é calculada conforme apresentada na equação 9 abaixo.

$$N_{P_i} = \sum_{v \in V_0} p_{i0}^v + \sum_{v \in V_1} p_{i1}^v \quad (9)$$

$$M_{P_i} = \sum_{l=1}^r p_i^v \quad (10)$$

[058] Da mesma forma, a variância de X para ambos os casos é dada como

$$Var_{N_{P_i}} = \sum_{v \in V_0} p_{i0}^v(1 - p_{i0}^v) + \sum_{v \in V_1} p_{i1}^v(1 - p_{i1}^v) \quad (11)$$

$$Var_{M_{P_i}} = \sum_{v \in V} p_i^v(1 - p_i^v) \quad (12)$$

[059] De acordo com a presente invenção, o aviso precoce para a parte  $P_i$  é comunicado quando  $N_{P_i} - M_{P_i} > \tau_{P_i}$ , em que  $\tau_{P_i}$  é decidido com base em  $N_{P_i}$  e  $Var_{N_{P_i}}$ .

[060] Uma vez detectado um aviso precoce para a parte  $P_i$  são determinadas as causas subjacentes possíveis na etapa 208 identificando e caracterizando um subconjunto das máquinas conectadas por regras que apontam eventuais causas para a anomalia. Em alguns casos, um subconjunto que se desvia significativamente do resto da população de máquinas conectadas em termos da taxa de falhas é pequeno em comparação com a população total, o que resulta em um caso de não aviso precoce quando é realizada uma análise a nível da população total. O sistema 100 da presente invenção considera dois cenários possíveis - 1) análise das causas subjacentes para encontrar um subconjunto das máquinas conectadas, quando é identificado um aviso precoce para alguma parte a nível da população total e 2) determinação de regras para encontrar um subconjunto de veículos, quando não é visível nenhum aviso precoce a nível da população total, quer dizer quando a diferença entre  $M_{P_i}$  e  $N_{P_i}$  é insignificante a nível da população, mas existe um pequeno subconjunto de população

desconhecido que apresenta um desvio significativo em relação ao resto da população em termos de taxa de falhas.

[061] De acordo com a presente invenção, para encontrar um subconjunto de veículos que possa ser a eventual razão para um aviso precoce já identificado, as probabilidades processadas por computador nas equações 7 e 8 acima referidas são usadas para determinar os momentos de falha esperados a nível dos veículos, quer dizer,  $e_i^v$ , o momento de falha esperado da parte  $P_i$  é processado por computador considerando que sobreviveu até ao momento  $t_0$  para cada  $v \in V$ , o que é dado pela equação 13 abaixo.

$$e_i^v = t_v' + \frac{1}{S(t_v)} \int_{t_v}^{\infty} S(t) \quad (13)$$

em que  $t_v$  e  $t_v'$  representam o tempo do veículo até ao início de  $t_0$  a partir de algum momento inicial e  $S(t)$  é a probabilidade de sobrevivência.

[062] A probabilidade de falha da parte  $P_i$  para cada  $v \in V$  é analisada usando o modelo de fiabilidade básica (BRM) convencional assim como o modelo de fiabilidade aumentado por sensores (SARM) que incorpora DTCs. No BRM, os parâmetros de falha são usados como parâmetros de escala e de forma na equação 13 para calcular  $e_{i1}^v$  para cada  $v \in V$ . Em uma forma de realização, o momento de venda do veículo  $v$  é usado como momento inicial, quer dizer  $t_v' = t_v = t_0 - t_s^v$ .

[063] No SARM, o tempo de falha esperado  $e_{i2}^v$  da parte  $P_i$  é determinado dividindo  $r$  veículos nos conjuntos  $V_0$  e  $V_1$ . Para cada  $v \in V_0$  é calculado o momento de falha esperado. Mas para os veículos em  $V_1$ , os parâmetros de retardamento são



usados na equação 13 para calcular o momento de falha esperado de  $P_i$ . Assim, o momento de ocorrência  $t_{D_j}$  de DTC  $D_j$  é usado como momento inicial, quer dizer  $t_v = t_0 - t_{D_j}$  e  $t'_v = t_{D_j} - t_s^v$ .

[064] Assim para cada  $v \in V$  existem dois momentos de falha esperados  $e_{i_1}^v$  e  $e_{i_2}^v$  da parte  $P_i$ , calculados usando o BRM e o SARM conforme acima descritos. Além disso,  $e_{i_1}^v$  e  $e_{i_2}^v$  são usados para definir uma pontuação de anomalia da parte  $P_i$  para  $v \in V$ , que é dada pela equação seguinte.

$$a_i^v = |e_{i_1}^v - e_{i_2}^v| \quad (14)$$

[065] Os veículos da população que apresentam uma pontuação de anomalia superior a um limiar predeterminado são selecionados para encontrar um subconjunto de veículos que possam apontar para as eventuais razões para o aviso precoce. Usando uma análise característica-a-característica, uma recolha de características, por exemplo, modelo, ano de fabricação, instalação de fabricação, geografia, fornecedor, etc. das máquinas conectadas que diferem estatisticamente entre os dois conjuntos, quer dizer, a totalidade da população e o subconjunto exibindo pontuações de anomalia elevadas são determinados.

[066] É igualmente possível que apenas um subconjunto reduzido e desconhecido de veículos apresente um desvio significativo do resto da população em termos de taxa de falhas e que a diferença entre  $N_{P_i}$  e  $M_{P_i}$  permaneça insignificante a nível da população. Para assegurar que estas situações são detectadas tem de ser identificado um subgrupo da população apresentando uma pontuação de anomalia elevada.

De acordo com uma forma de realização é usada uma técnica de descoberta de subgrupo, em que são agregadas regras de associação para se obter um pequeno conjunto de regras que cobrem uma fracção grande dos dados. Conforme descrito acima, a cada máquina conectada ou parte desta pode ser atribuída uma pontuação de anomalia e, de acordo com a presente invenção, podem ser identificados subgrupos de veículos com pontuações de anomalia elevadas, caracterizadas por regras com base nos atributos ou nas características da máquina/parte, tais como o modelo, o ano de fabricação, a instalação de fabricação, a geografia, o fornecedor, etc.. Cada máquina conectada pode ser visualizada como um registro com as respectivas características em campos e uma pontuação de anomalia para a parte  $P_i$  que é discretizada em dois níveis, "elevada" ou "normal", quer dizer, o campo de anomalia divide o conjunto de veículos  $V$  em dois conjuntos designados por  $E_i$  representando um conjunto de aviso precoce e  $NE_i$  representando um conjunto de não aviso precoce para a parte  $P_i$ , em que

$$E_i = \{v \in V : a_i^v > \tau_a\} \quad (15)$$

e  $NE_i = E_i^c$  quer dizer complementam o conjunto de  $E_i$ .

[067] De acordo com uma forma de realização, as regras de associação com uma pontuação de anomalia "elevada"  $E_i$  primeiro são minerados como o consequente, por ex.  $X \& Y \dots \rightarrow E_i$ , que satisfazem igualmente uma confiança mínima razoável (por ex.  $\text{conf}(r) > 0.75$ ). Subsequentemente estas regras são ordenadas por ordem decrescente de suporte para escolher um subconjunto principal que cobre uma fracção suficientemente grande dos dados (por ex. pelo menos 50%). Em seguida as regras são agregadas usando uma técnica com base na densidade

tal como DBScan e uma medida de distância que é inversamente proporcional ao grau de sobreposição entre duas regras, quer dizer, o número de registros que satisfazem ambas as regras. Como resultado, cada agregado de regras contém regras que se sobrepõem fortemente umas às outras; ao contrário, as regras de agregados diferentes apresentam uma sobreposição mútua reduzida. Finalmente, são selecionadas uma ou mais regras de cada agregado que apresentam um suporte e uma confiança elevados, gerando um conjunto de regras reduzido que identifica um subconjunto de máquinas conectadas com pontuações de anomalia predominantemente elevadas; os antecedentes destas regras apontam eventuais causas para as pontuações de anomalia elevadas observadas. Além disso, cada regra pode ser refinada pelas respectivas "exceções" por re-execução do procedimento acima referido não apenas dos dados cobertos pelos antecedentes da regra, mas usando o  $NE_i$  como consequente. O procedimento de determinação de regras, conforme acima descrito, tem de ser executado independentemente de uma anomalia ser detectada a nível da população total ou não, considerando que é concebido particularmente para a situação em que a anomalia não é visível a nível da população.

#### RESULTADOS EXPERIMENTAIS

[068] Os sistemas e os métodos da presente invenção foram validados pela previsão de falhas futuras de um conjunto de dados reais de um fabricante de veículos automóveis e pela comparação desta com falhas reais e falhas esperadas calculadas usando o BRM. Foram considerados três cenários encontrados nos dados do mundo real: 1) quando os volumes de falha reais são significativamente superiores ao

esperado, quer dizer, um caso de aviso precoce; 2) quando os volumes de falha reais são inferiores ao esperado, e; 3) quando os volumes de falha reais e esperados coincidem. Os volumes de falha "esperados" se referem àqueles previstos por um modelo de fiabilidade básica (BRM) convencional apenas com base em dados relativos a falhas passadas. Em cada um dos casos acima referidos se verificou que o modelo aumentado (SARM) da presente invenção prevê volumes futuros que estão mais próximos dos números reais do que o modelo convencional (BRM). O modelo SARM da presente invenção foi igualmente validado por análise das causas subjacentes e determinação de regras com base em dados sintéticos.

[069] Dados reais: foram considerados três conjuntos de dados incluindo dados relativos à venda, dados relativos ao DTC e dados relativos a reclamações. A Tabela I apresenta o número de veículos únicos ou os números de identificação de veículo VINS e o período de tempo para o qual os três conjuntos de dados estão disponíveis.

TABELA I - CARACTERIZAÇÃO DE DADOS DO CONJUNTO DE

DADOS

Conjunto de dados	Dados relativos à venda	Dados relativos ao DTC	Dados relativos a reclamações
Número de VINS únicos	1,9M	54K	0,5M
Período de tempo	3 anos	2 anos	2 anos

[070] Com base na disponibilidade de dados, são escolhidas duas partes  $P_G$  e  $P_B$  para validar os métodos e os sistemas da presente invenção e os resultados foram processados por computador para os três cenários acima referidos. A caracterização de dados para estas duas partes é dada na Tabela II.

TABELA II - CARACTERIZAÇÃO DE DADOS DAS PARTES  $P_G$  E $P_B$ 

Parte escolhida	$P_G$	$P_B$
VINs únicos nos dados relativos a reclamações	513	7651
VINs únicos nos dados relativos ao DTC	445	1239
Período de dados relativos a reclamações	1,8 anos	2 anos
Período de dados relativos ao DTC	6 meses	9 meses

[071] De acordo com os métodos da presente invenção primeiro é identificado um conjunto de DTCs discriminativos para cada parte. A Tabela III apresenta as regras identificadas para as partes  $P_G$  e  $P_B$ .

TABELA III - REGRAS DOS DTCS IDENTIFICADOS PARA AS PARTES  $P_G$  E  $P_B$ 

Parte	Regra n.º	Antecedente	Consequente	Suporte em D	Suporte em F	Confiança	Cobertura
$P_G$	1	P2162	$P_G$	0,168	0,485	0,90	83%
	2	P07E7	$P_G$	0,15	0,43	0,73	
	3	P07E0	$P_G$	0,17	0,42	0,87	
$P_B$	1	B1304	$P_B$	0,18	0,58	0,93	84%
	2	B100D	PB	0,17	0,55	0,83	
	3	B1D21	PB	0,15	0,48	0,87	

[072] É evidente que existem três regras com  $\tau_p = 0.7$  e  $R_c = 0.8$  para ambas as partes. É igualmente evidente que o suporte de cada regra na base de dados geral D e em um conjunto de falha  $D_f$ . Assim, o  $\text{disc}_i$  para a parte  $P_G$  é {P2162, P07E7, P07E0} e para a parte  $P_B$  é {B1304, B100D, B1D21}. Para cada parte, os parâmetros de retardamento e de falha de cada DTC em  $\text{disc}_i$  são determinados usando um modelo gráfico Bayesiano conforme acima explicado com referência à etapa 204 do método da presente invenção. Para a experiência foi usada uma biblioteca python "pymc" para estimar os parâmetros do

modelo Bayesiano por meio da amostragem do tipo Markov Chain Monte Carlo (MCMC).

[073] Conforme acima descrito no modelo Bayesiano se parte do princípio que o retardamento entre a falha da parte e a ocorrência do DTC segue uma distribuição de Weibull com parâmetros de retardamento como parâmetros de escala e de forma da distribuição. Foi confirmado empiricamente que a distribuição de Weibull é adequada para a variável  $\tau_{ij}$ . A adequação da distribuição de Weibull foi comparada com a distribuição de Gaussian para a variável  $\tau_{ij}$ . A FIG. 4 ilustra gráficos (quantil-quantil) QQ entre os dados reais da variável  $\tau_{ij}$  e os dados de amostra das distribuições de Weibull e de Gaussian, com parâmetros estimados a partir do modelo Bayesiano, para os seis DTCs. É claramente evidente que a distribuição de Weibull é mais adequada do que a distribuição de Gaussian para a variável  $\tau_{ij}$ . Além dos gráficos QQ foi usado um teste do tipo Kolmogorov-Smirnov(K-S) para verificar a adequação para a variável  $\tau_{ij}$ . A Tabela IV apresenta o valor p da distribuição de Weibull e de Gaussian para a variável  $\tau_{ij}$ . São igualmente evidente os valores dos parâmetros de retardamento e de falha em termos de dias calculados a partir do momento de venda do veículo, que foram determinados para partes e os respectivos DTCs discriminativos usando o modelo Bayesiano.

TABELA IV - PARÂMETROS DETERMINADOS USANDO O MODELO GRÁFICO BAYESIANO

Parte	DTC	$\alpha_{Dj}$	$\beta_{Dj}$	$\alpha_i$	$\beta_i$	Valor p usando Weibull	Valor p usando Gaussian
P <sub>G</sub>	P2162	9,572	0,81	3603,53	3,58	0,44	0,31
	P07E7	32,637	1,65	3603,53	3,58	0,39	0,21
	P07E0	29,723	1,71	3603,53	3,58	0,31	0,19

P <sub>B</sub>	B1304	12,04	0,52	900,45	2,24	0,37	0,12
	B100D	28,97	1,32	900,58	2,24	0,34	0,14
	B1D21	34,23	1,45	900,58	2,24	0,31	0,12

[074] De acordo com a presente invenção, o número futuro de falhas das partes P<sub>G</sub> e P<sub>B</sub> é previsto usando os parâmetros de retardamento e de falha apresentados na Tabela IV. No caso da parte P<sub>G</sub>, os dados relativos a reclamações e os dados relativos ao DTC até t<sub>0</sub> são usados para identificar um conjunto de DTCs discriminativos e para determinar os parâmetros do modelo Bayesiano. O número de falhas para o mês M1 (> t<sub>0</sub>) é previsto e comparado com o número de falhas esperado e as falhas reais. Da mesma forma, no caso da parte P<sub>B</sub>, é previsto o número de falhas para dois meses M2 e M3. A FIG. 5 apresenta o número de falhas previsto (usando SARM) vs real vs esperado (por BRM) para os três cenários acima referidos, demonstrando claramente o melhor desempenho preditivo do SARM da presente invenção em comparação com as falhas esperadas usando o BRM. É evidente que as previsões apontam para um aviso precoce para a parte P<sub>G</sub>, considerando que a diferença entre os valores previstos (SARM) e esperados (BRM) é significativa. Além disso, as previsões SARM coincidem melhor com as falhas observadas em M1.

[075] A FIG. 6 apresenta a tendência semanal de esperado vs real vs previsto, para a parte P<sub>G</sub>. É claramente evidente que existe um caso de aviso precoce na semana 6 e uma tendência de aviso precoce similar ocorre nas semanas futuras (> t<sub>0</sub>) e que o modelo SARM da presente invenção prevê corretamente esta tendência.

[076] Dados sintéticos: Foram simulados três cenários: 1) NW: sem aviso precoce; 2) EW-1: aviso precoce-1,

e; 3) EW-2: aviso precoce-2. No caso de NW, foram gerados 0,1M registros para veículos com data de compra, quilômetros percorridos, datas de ocorrência de DTC e datas de falha de partes; assumindo que as datas de compra estão distribuídas uniformemente por um período de 5 anos. O número de quilômetros percorridos por dia para cada veículo foi determinado a partir de uma distribuição uniforme. Os momentos de falha foram gerados para 10 partes usando distribuições de Weibull com diferentes parâmetros de escala e de forma. Foram igualmente gerados os dados relativos ao DTC com códigos DTC, que se tornam observáveis antes da falha da parte. O retardamento entre a falha da parte e a ocorrência do DTC foi gerado usando a distribuição de Weibull com parâmetros diferentes para cada DTC. Os atributos ou as características dos veículos gerados são apresentados na Tabela V.

TABELA V - ATRIBUTOS E CARACTERÍSTICAS DE VEÍCULOS EM DADOS SINTÉTICOS

Atributo	Intervalo
Ano do modelo	2010 até 2014
Instalação de fabricação	1, 2, 3
Geografia	Estados Unidos, China, Reino Unido, Índia
Portas	4 e 5

[077] No caso de EW-1, o aviso precoce em veículos do ano do modelo 2012 foi induzido por antecipação dos momentos de falha da parte  $P_i$  nestes veículos. Da mesma forma, no caso de EW-2, o aviso precoce em veículos do ano do modelo "2012", da geografia "EUA", e fabricados na instalação 1 e 2 foi induzido por antecipação dos momentos de falha da parte  $P_i$  nestes veículos. Considerando o momento  $t_0$  as falhas



da parte  $P_i$  foram previstas usando uma abordagem de aviso precoce e comparadas com as falhas esperadas e com as falhas reais.

[078] A Tabela VI contém as pontuações de anomalias médias para veículos de toda a população para cada um dos três casos acima referidos.

TABELA VI - ATRIBUTOS OU CARACTERÍSTICAS DE VEÍCULOS EM DADOS SIMULADOS

	NW	EW-1	EW-2
Pontuação de anomalia	6,573	11,573	7,754

[079] É evidente que a pontuação de anomalia para EW-1 é quase o dobro da pontuação do caso sem aviso (NW) e, por conseguinte, visível a nível da população total. Contudo, o aviso de EW-2 não é visível na população total, sendo que a respectiva pontuação de anomalia é equivalente àquela de NW ou do caso sem aviso precoce.

[080] A análise das causas subjacentes foi realizada para revelar a eventual causa de EW-1. EW-2 pode ser detectado apenas por determinação de regras considerando que é invisível com base nas pontuações de anomalias médias a um nível da população total. A análise das causas subjacentes para a parte  $P_i$  é demonstrada por meio de um suporte de ferramentas analíticas visuais ilustrado na FIG. 7, que apresenta um mapa de calor entre o mês de falhas esperado da parte  $P_i$  para os veículos, calculado usando o BRM (eixo x) e o SARM (eixo y). É igualmente evidente a distribuição de vários atributos de veículos selecionados para observação. A região muito abaixo da diagonal do mapa de calor contém veículos em que o SARM prevê uma falha muito mais precoce do

que o BRM, quer dizer, os veículos envolvidos no aviso precoce. A seleção de uma gama desta região atualiza as distribuições por atributos para aquelas para este conjunto restrito. Usando as distribuições antes e depois de uma seleção desta natureza são comparadas, revelando a razão para o aviso precoce. A FIG. 7 sugere claramente que depois de selecionar veículos na região direita inferior, quer dizer aqueles com pontuações de anomalia elevadas, a probabilidade do atributo "ano do modelo 2012" aumenta significativamente, o que demonstra que uma percentagem elevada de veículos que se situam em uma região de pontuações de anomalia elevada são carros do ano do modelo 2012.

[081] Para a determinação de regras da parte Pi no caso EW-2, os veículos em que a parte Pi não falhou até  $t_0$  são divididos em dois conjuntos: "aviso precoce" e "aviso não precoce", usando um limiar  $\tau_a$  como a média da pontuação de anomalia para todos os veículos. Além disso, as regras e as exceções dos atributos dos veículos são identificadas usando técnicas, conforme explicadas com referência às equações 13, 14 e 15. A FIG. 8 ilustra um par de exceção à regra exemplificativo descoberto para um conjunto de aviso precoce, em que são apresentados a confiança, o suporte e a elevação para a regra; além disso uma exceção à regra é apresentada de forma similar. Isto demonstra que os veículos do ano do modelo 2012 e da geografia (EUA) e fabricados em todas as instalações apresentam um aviso precoce, com uma confiança de 82% e um suporte de 15,6%, com uma exceção: nos carros cobertos por esta regra, aqueles fabricados na instalação b situam-se mais frequentemente no conjunto de aviso não precoce do que outros.

[082] Os sistemas e os métodos da presente invenção providenciam uma estrutura completa para prever avisos precoces de volumes de falha inesperados usando DTCs com base em sensores conjuntamente com dados relativos à falha de partes e combinam os avisos precoces de falha com uma análise das causas subjacentes assim como uma descoberta de subgrupo. Os sistemas e os métodos da presente invenção exploram a disponibilidade de sensores indicadores incorporados em veículos modernos para sinalizar avisos precoces desta natureza por comparação das previsões de fiabilidade convencional com aquelas de um modelo aumentado com informação de sensores conforme recolhida e transmitida por meio da "internet industrial" de populações de máquinas conectadas altamente. Quando os dois modelos diferem significativamente é um indicativo de uma anomalia a nível da população e eventualmente de um indicador da eventual necessidade de uma repetição no futuro. Além disso, para detectar anomalias da forma acima referida, os sistemas e os métodos da presente invenção permitem expandir para descobrir eventuais causas subjacentes à anomalia, de forma que possa ser resolvida antecipadamente. Finalmente, os sistemas e os métodos da presente invenção identificam anomalias em pequenos subgrupos que não são visíveis estatisticamente na população geral. A eficácia dos sistemas e dos métodos da presente invenção é evidente a partir das experiências realizadas usando dados da vida real em que foram detectados avisos precoces reais e experiências usando dados gerados sinteticamente em que foram corretamente estabelecidas uma análise das causas subjacentes e uma descoberta de regras.

[083] Embora os sistemas e os métodos da

presente invenção tenham sido ilustrados com referência à previsão de falhas na indústria automóvel por meio dos dados experimentais acima referidos para o perito na técnica é evidente que a previsão de falhas desempenha um papel importante em vários domínios e os sistemas e os métodos da presente invenção podem ser aplicados aos domínios do software, da saúde e dos seguros e similares.

[084] A descrição escrita descreve o objeto da presente invenção para permitir ao perito na técnica realizar e usar as formas de realização da invenção. O âmbito das formas de realização do objeto definido no presente documento pode incluir outras modificações que ocorram ao perito na técnica. Estas modificações pretendem fazer parte do âmbito quando apresentam elementos similares que não diferem da linguagem literal das reivindicações ou quando incluem elementos equivalentes com diferenças não substanciais da linguagem literal.

[085] Contudo, é evidente que o âmbito da proteção se estende a um programa desta natureza e além disso a um meio legível por computador apresentando uma mensagem neste; este meio de armazenamento legível por computador contém meios de código de programa para implementar uma ou mais etapas do método, quando o programa é executado em um servidor ou um dispositivo móvel ou qualquer outro dispositivo programável adequado. O dispositivo de hardware pode ser qualquer tipo de dispositivo que possa ser programado incluindo por ex. qualquer tipo de computador como um servidor ou um computador pessoal, ou similar, ou qualquer combinação destes. O dispositivo pode igualmente incluir meios que podem ser por ex. meios de hardware tais como por

ex. circuitos integrados específicos de aplicação (ASIC), redes de portas lógicas programáveis (FPGA) ou uma combinação de meios de hardware e de software, por ex. um ASIC e uma FPGA, ou pelo menos um microprocessador e pelo menos uma memória com módulos de software localizados nesta. Assim, os meios podem incluir meios de hardware e de software. As formas de realização do método descritas no presente documento podem ser implementadas em hardware e em software. O dispositivo pode igualmente incluir meios de software. Alternativamente, a invenção pode ser implementada em diferentes dispositivos de hardware, por ex. usando uma pluralidade de CPUs.

[086] As formas de realização no presente documento podem compreender hardware e software. As formas de realização que são implementadas em software incluem mas não estão limitadas a *firmware*, software residente, microcódigos, etc.. As funções desempenhadas pelos vários módulos que constituem o sistema da presente invenção e são descritos no presente documento podem ser implementados em outros módulos ou em combinações de outros módulos. Para efeitos desta descrição, um meio utilizável por computador ou legível por computador pode ser qualquer aparelho que compreenda, armazene, comunique, propague ou transporte o programa para o uso por ou em conexão com o sistema, aparelho ou dispositivo de execução de instruções. Os vários módulos descritos no presente documento podem ser implementados como módulos de software e/ou hardware e podem ser armazenados em qualquer tipo de meio legível por computador não transitório ou outro dispositivo de armazenamento. Alguns exemplos não limitativos de meios legíveis por computador não transitórios incluem

CDs, DVDs, BLU-RAY, memórias *flash* e unidades de disco rígido.

[087] Um sistema de processamento de dados adequado para o armazenamento e/ou a execução de códigos de programa inclui pelo menos um processador acoplado diretamente ou indiretamente a elementos de memória por meio de um sistema de barramento. Os elementos de memória podem incluir uma memória local usada durante a execução real do código de programa, memória de grande capacidade e memórias em cache que providenciam um armazenamento temporário de pelo menos algum código de programa para reduzir o número de vezes que o código tem de ser consultado a partir da memória de grande capacidade durante a execução.

[088] Além disso, embora as etapas de processo, as etapas de método, as técnicas ou similares possam ser descritos em uma ordem sequencial, estes processos, métodos e técnicas podem ser configurados para trabalhar em ordens alternadas. Por outras palavras, qualquer sequência ou ordem de etapas que pode ser descrita não indica necessariamente que as etapas tenham de ser realizadas nessa ordem. As etapas dos processos descritas no presente documento podem ser realizadas em qualquer ordem prática. Além disso, algumas etapas podem ser realizadas simultaneamente.

[089] A descrição precedente foi apresentada com referência a várias formas de realização. Para os peritos na técnica é evidente que podem ser realizadas alterações e mudanças nas estruturas e nos métodos de operação descritos sem abandonar o princípio, o espírito e o âmbito.

REIVINDICAÇÕES

1. MÉTODO IMPLEMENTADO POR COMPUTADOR (200) PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS, caracterizado pelo método (200) compreender:

identificar conjuntos de Códigos de Diagnóstico de Falhas (DTCs) (114) discriminativos a partir de DTCs gerados por falhas passadas, em que os conjuntos de DTCs discriminativos correspondem a partes predefinidas associadas das máquinas conectadas;

gerar um modelo de dependência condicional temporal com base na dependência temporal entre a falha das partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados; e

prever falhas futuras com base na dependência condicional temporal gerada e na ocorrência e na não-ocorrência de DTCs; e

realizar uma análise das causas subjacentes às falhas futuras previstas para a mineração de fiabilidade previsível, em que a etapa de realizar uma análise das causas subjacentes compreende identificar um subconjunto da população com uma eventual anomalia quando não é gerado um aviso precoce a nível da população, em que a etapa de realizar a análise das causas subjacentes compreende:

(i) definir uma pontuação de anomalia para cada uma das máquinas conectadas;

(ii) associar cada uma das máquinas conectadas a um registro compreendendo um conjunto de características predefinidas e a pontuação de anomalia definida;

(iii) discretizar a pontuação de anomalia definida em um nível elevado predeterminado ou em um nível normal;

(iv) realizar uma mineração de regras de associação para identificar regras com a pontuação de anomalia de nível elevado;

(v) agregar as regras de associação identificadas usando uma técnica com base na densidade para formar agregados de regras;

(vi) selecionar uma ou mais regras de cada um dos agregados de regras que apresentam suporte e confiança elevados; e

identificar o subconjunto da população com a pontuação de anomalia de nível elevado com base em antecedentes das uma ou mais regras selecionadas que são indicativas de eventuais razões para a pontuação de anomalia de nível elevado.

2. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 1, caracterizado pela etapa de identificar conjuntos de DTCs discriminativos ser baseada na mineração de regras de associação.

3. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 2, caracterizado pela mineração de regras de associação compreender o uso da técnica Apriori.

4. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 2, caracterizado pelos antecedentes das regras identificadas pela técnica de mineração de regras de associação formarem o conjunto de DTCs discriminativos.

5. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 1, caracterizado pelo modelo de



dependência condicional temporal ser uma rede Bayesiana (116).

6. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 1, caracterizado pela dependência temporal entre a falha das partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados seguir uma distribuição de Weibull.

7. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 1, caracterizado pela etapa de gerar um modelo de dependência condicional temporal ser seguida por uma etapa de separar a população de máquinas conectadas em um primeiro conjunto compreendendo máquinas conectadas em que os DTCs não são gerados em um período de tempo predeterminado e em um segundo conjunto compreendendo máquinas conectadas em que pelo menos um dos DTCs é gerado no período de tempo predeterminado.

8. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 7, caracterizado pela etapa de prever falhas futuras compreender a etapa de processar por computador a probabilidade de falha com base em ambas, na ocorrência e na não-ocorrência de DTCs, na população separada de máquinas conectadas e de gerar um aviso precoce quando o número previsto de falhas é superior ao número esperado de falhas com base nos dados relativos a falhas passadas por um valor predefinido.

9. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 8, caracterizado pelo valor predefinido ser baseado no número previsto de falhas e a variância de uma variável aleatória representando o número de

falhas das partes predefinidas em um período de tempo predeterminado.

10. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 1, caracterizado pela etapa de realizar uma análise das causas subjacentes compreender identificar o subconjunto da população quando é gerado um aviso precoce para pelo menos uma das partes predefinidas.

11. MÉTODO IMPLEMENTADO POR COMPUTADOR (200), de acordo com a reivindicação 7, caracterizado pela etapa de realizar uma análise das causas subjacentes quando é gerado um aviso precoce para pelo menos uma das partes predefinidas compreender:

(i) calcular um primeiro momento de falha esperado com base nos dados relativos às falhas passadas;

(ii) calcular um segundo momento de falha esperado com base no momento de falha esperado do segundo conjunto da população separada com base ainda em parâmetros retardadores predefinidos e no momento de ocorrência de pelo menos um DTC;

(iii) definir uma pontuação de anomalia para cada uma das máquinas conectadas com base no primeiro momento de falha esperado e no segundo momento de falha esperado calculados;

(iv) realizar iterativamente as etapas desde (i) até (iv) para características predefinidas das máquinas conectadas; e

(v) identificar o subconjunto da população apresentando a pontuação de anomalia superior a um limiar predefinido, em que o subconjunto identificado indica eventuais razões para o aviso precoce para cada uma das características predefinidas.

12. SISTEMA (100) PARA A MINERAÇÃO DE FIABILIDADE PREVISÍVEL EM UMA POPULAÇÃO DE MÁQUINAS CONECTADAS, caracterizado pelo sistema (100) compreender:

um ou mais processadores;

um dispositivo de interface de comunicação;

um ou mais dispositivos de armazenamento de dados internos acoplados operacionalmente com o um ou mais processadores para o armazenamento:

um módulo de entrada (112) configurado para receber Códigos de Diagnóstico de Falhas (DTCs) (114) a partir de sistemas de diagnóstico de bordo (110) de partes predefinidas das máquinas conectadas;

um identificador de padrões de DTC configurado para identificar conjuntos de DTCs discriminativos a partir dos DTCs gerados por falhas precedentes, em que os conjuntos de DTCs discriminativos correspondem a partes predefinidas das máquinas conectadas;

um gerador de rede Bayesiana (116) configurado para gerar um modelo de dependência condicional temporal com base na dependência temporal entre a falha de partes predefinidas a partir de dados relativos a falhas passadas e os conjuntos de DTCs discriminativos identificados; e

um previsor de falhas (118) configurado para prever falhas futuras com base na dependência condicional temporal gerada e na ocorrência e na não-ocorrência de DTCs; e

um analisador (120) configurado para realizar uma análise das causas subjacentes às falhas futuras previstas para a mineração de fiabilidade previsível, em que a etapa de realizar uma análise das causas subjacentes compreende identificar um subconjunto da população com uma eventual

anomalia quando não é gerado um aviso precoce a nível da população, em que a etapa de realizar a análise das causas subjacentes compreende:

(i) definir uma pontuação de anomalia para cada uma das máquinas conectadas;

(ii) associar cada uma das máquinas conectadas a um registro compreendendo um conjunto de características predefinidas e a pontuação de anomalia definida;

(iii) discretizar a pontuação de anomalia definida em um nível elevado predeterminado ou em um nível normal;

(iv) realizar uma mineração de regras de associação para identificar regras com a pontuação de anomalia de nível elevado;

(v) agregar as regras de associação identificadas usando uma técnica com base na densidade para formar agregados de regras;

(vi) selecionar uma ou mais regras de cada um dos agregados de regras que apresentam suporte e confiança elevados; e

identificar o subconjunto da população com a pontuação de anomalia de nível elevado com base em antecedentes das uma ou mais regras selecionadas que são indicativas de eventuais razões para a pontuação de anomalia de nível elevado.

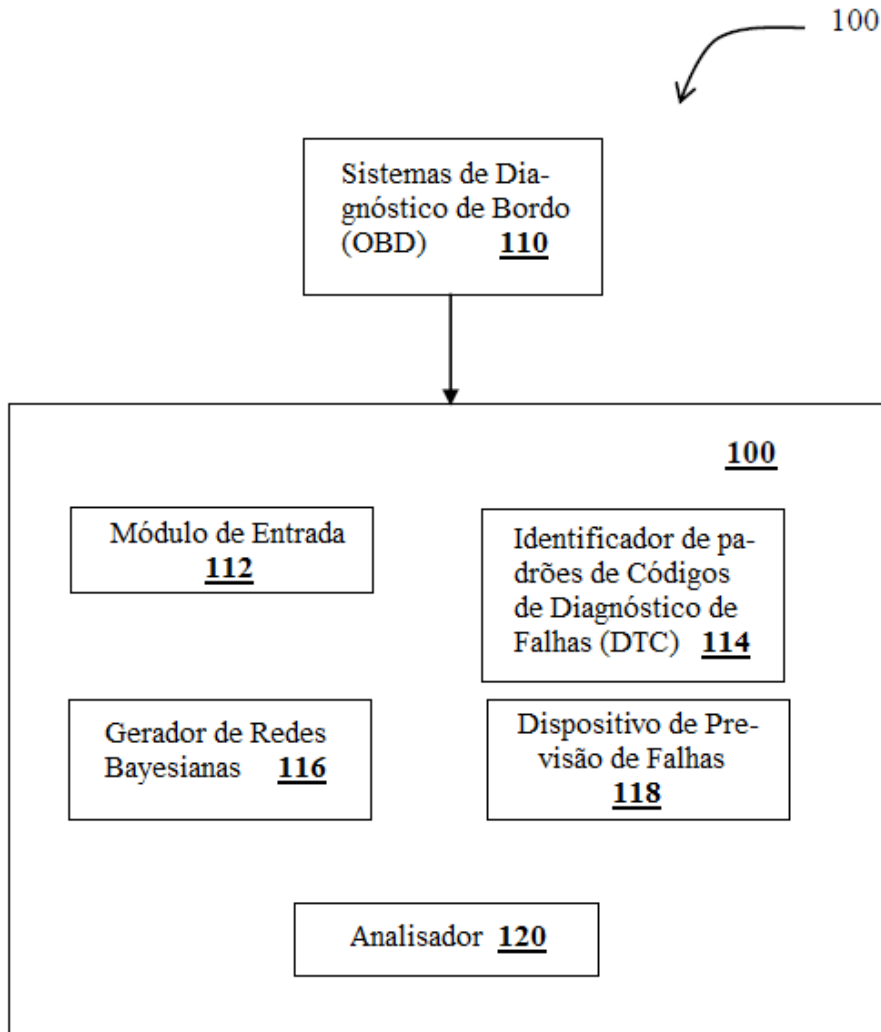


FIG. 1

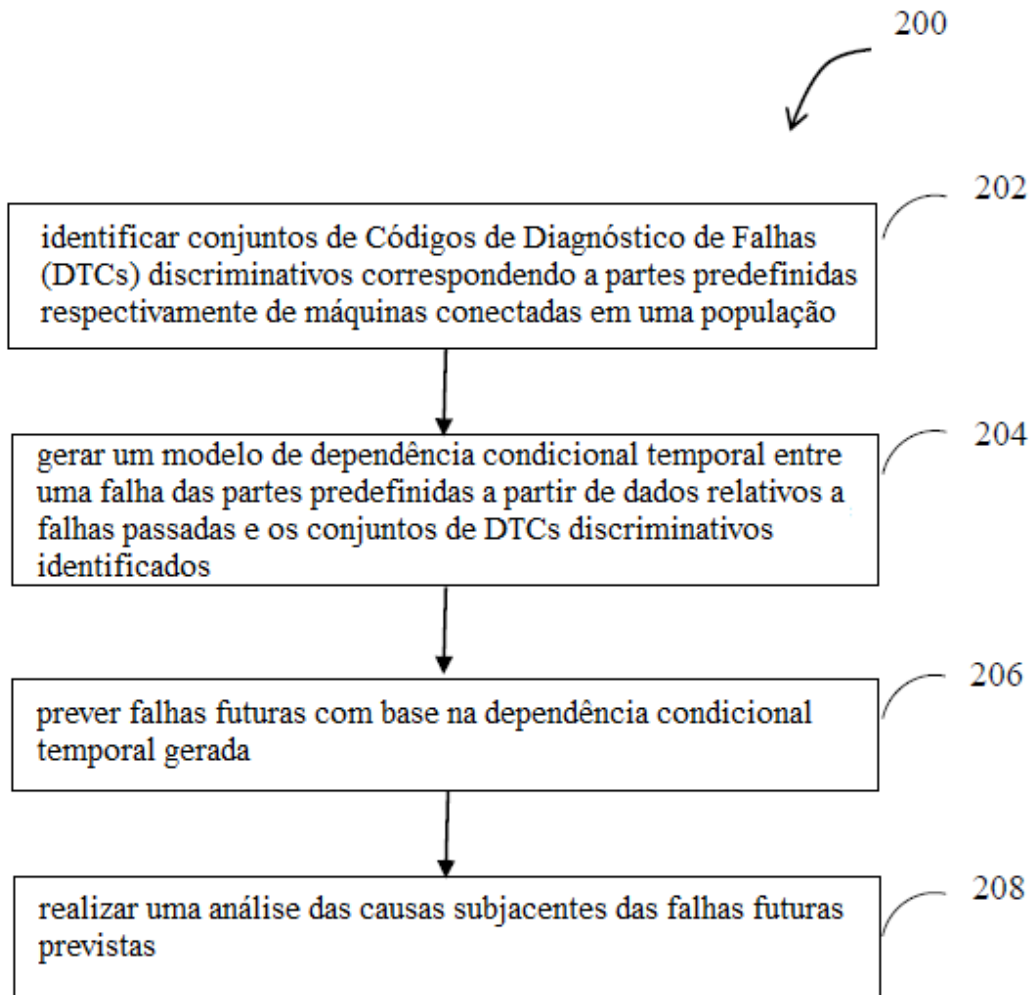


FIG. 2

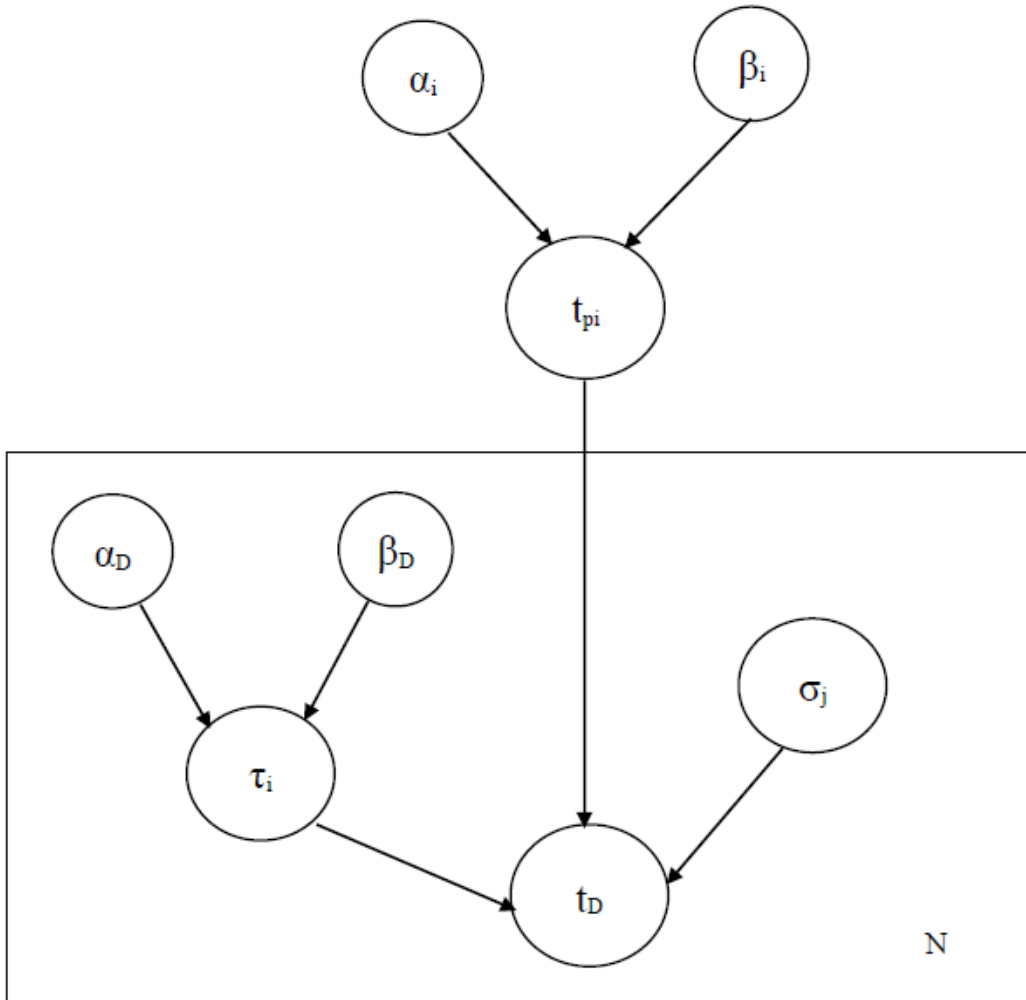


FIG. 3

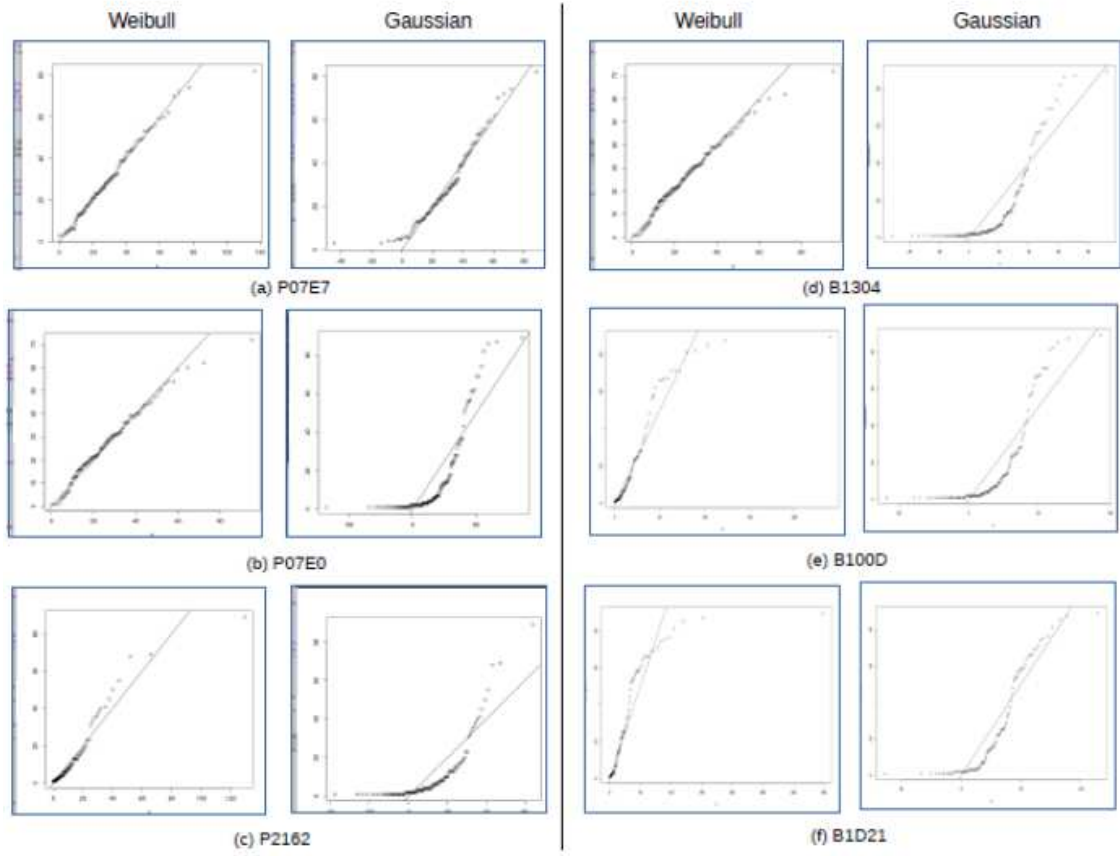


FIG. 4



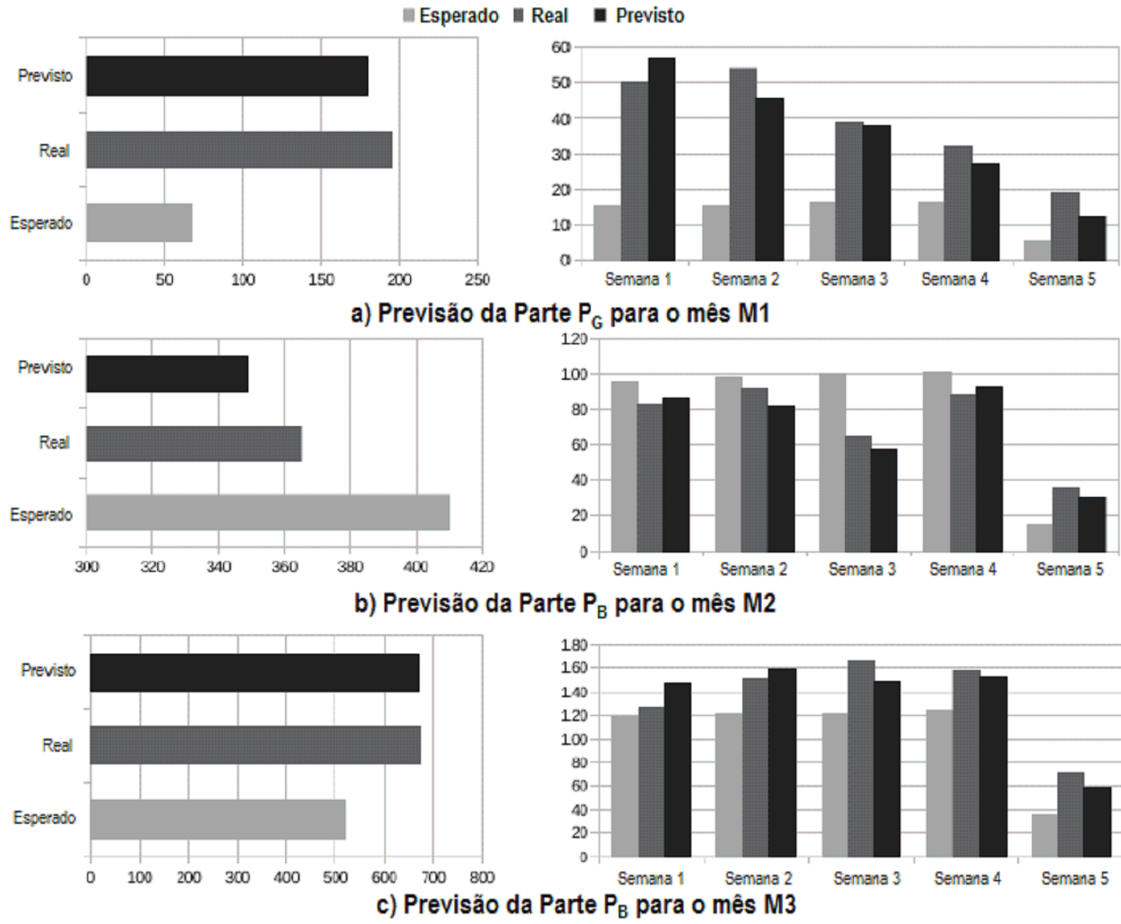


FIG. 5

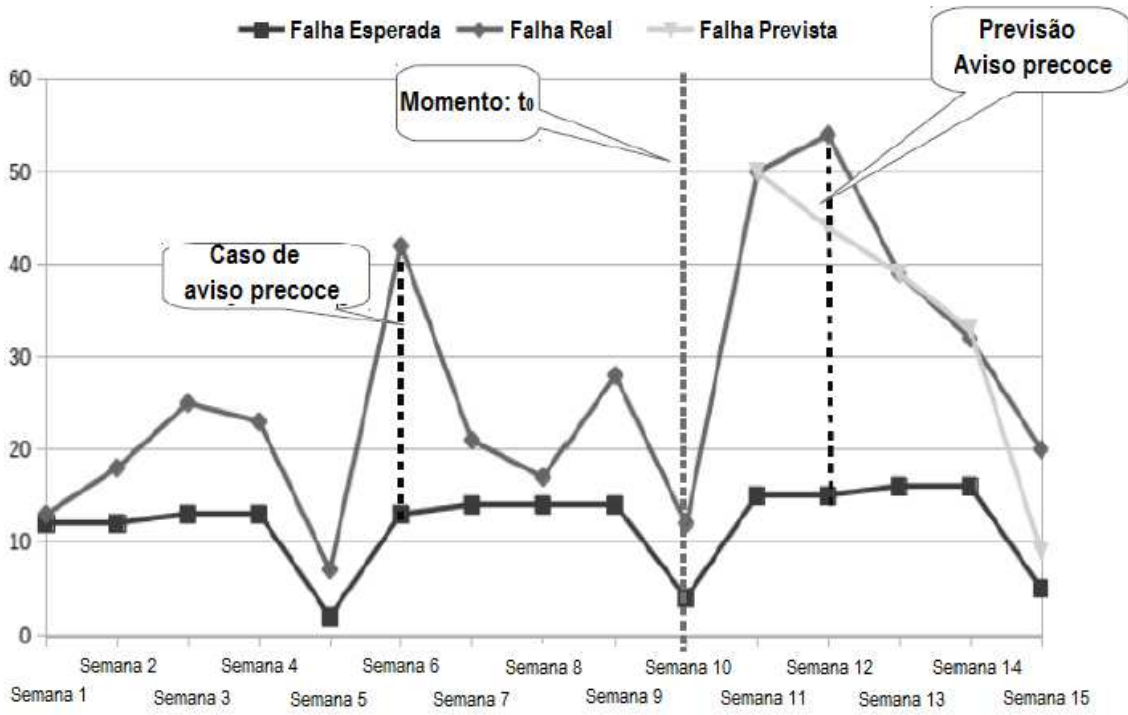


FIG. 6

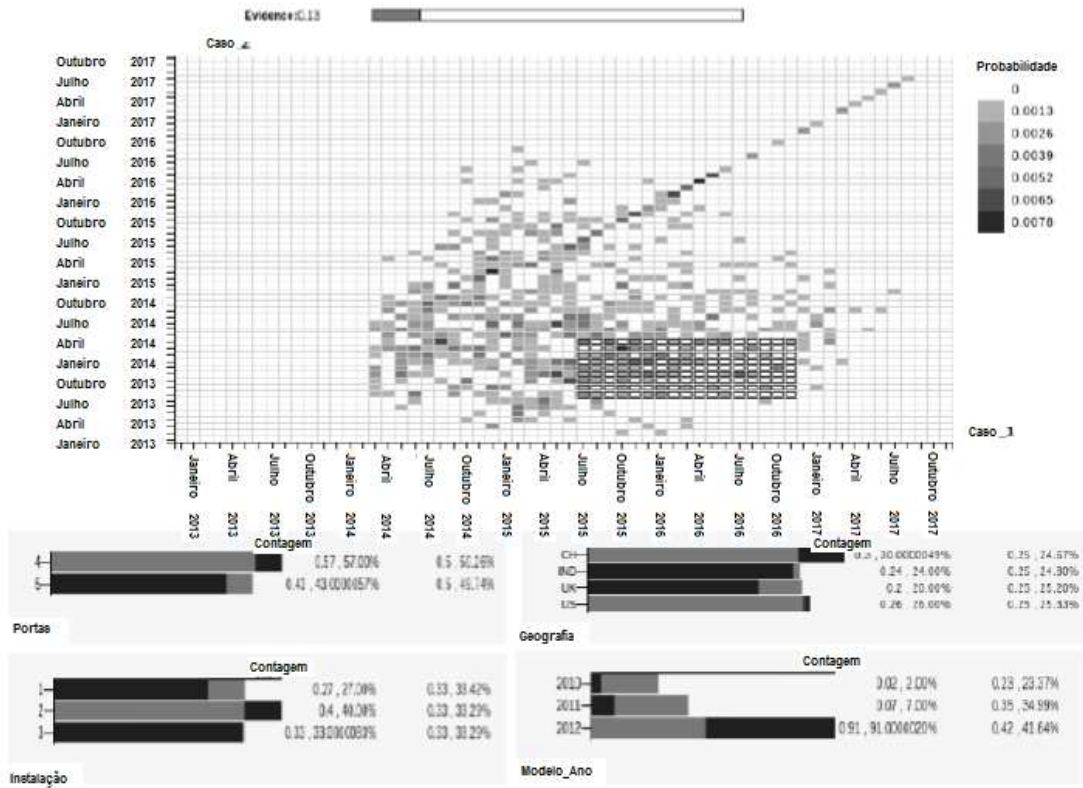


FIG. 7

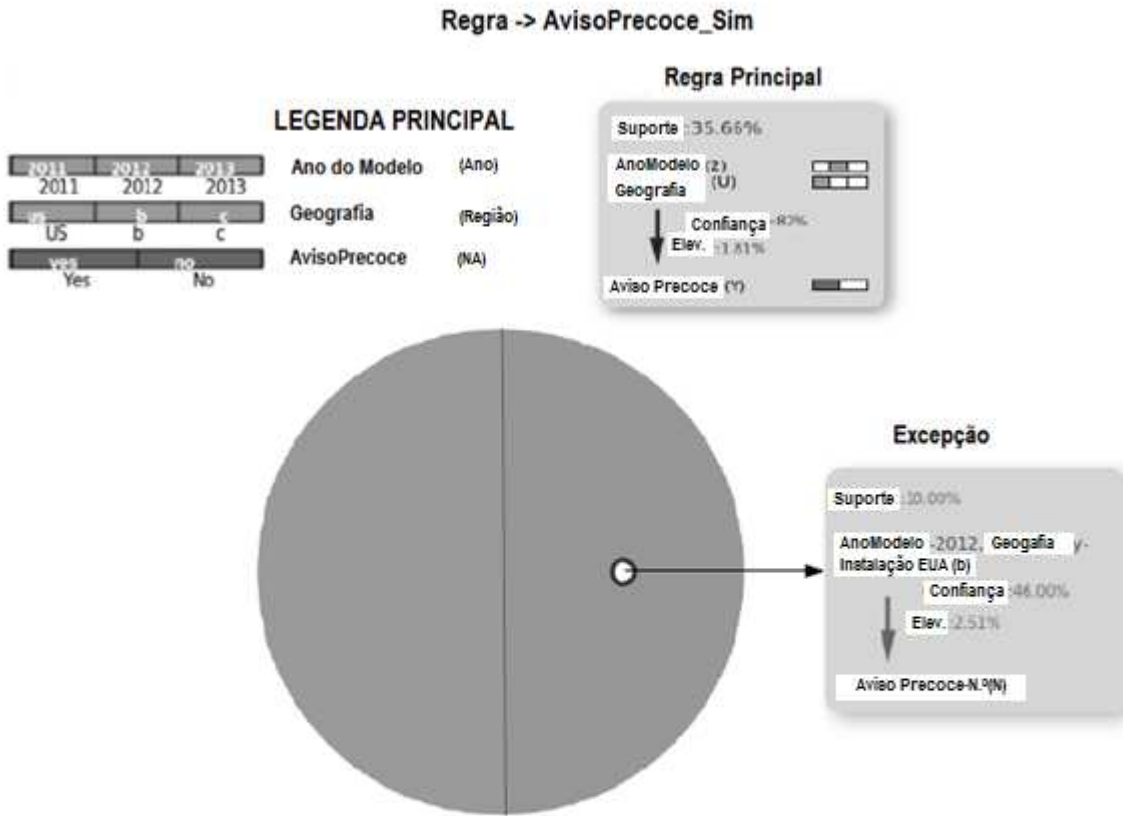


FIG. 8