(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103064987 A
(43)申请公布日 2013.04.24

(21)申请号 201310037691.8

(22)申请日 2013.01.31

(71)申请人 五八同城信息技术有限公司
地址 300457 天津市滨海新区第一大街 79 号泰达 MSD-C 区 -C3 座 2801 房间

(72)发明人 王永康 张爱华

(74)专利代理机构 工业和信息化部电子专利中心 11010
代理人 田俊峰

(51)Int.Cl.
G06F 17/30(2006.01)
G06Q 30/02(2012.01)

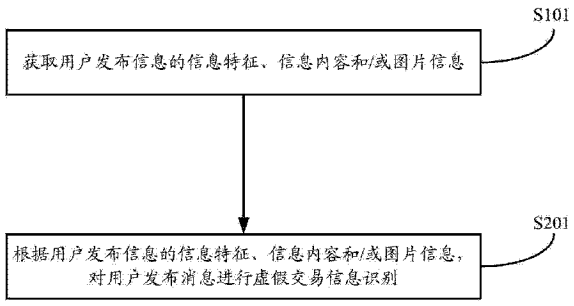
权利要求书2页 说明书6页 附图1页

(54)发明名称

一种虚假交易信息识别方法

(57)摘要

本发明公开了一种虚假交易信息识别方法，包括：步骤S101，获取用户发布信息的信息特征、信息内容和/或图片信息；步骤S201，根据用户发布信息的信息特征、信息内容和/或图片信息，对用户发布消息进行虚假交易信息识别。本发明可以大大的减少交易信息的虚假量，提高交易信息的真实性，增加用户体验，同时可以大大减少人力成本。



CN 103064987 A

1. 一种虚假交易信息识别方法，其特征在于，包括：

步骤 S101，获取用户发布信息的信息特征、信息内容和 / 或图片信息；

步骤 S201，根据用户发布信息的信息特征、信息内容和 / 或图片信息，对用户发布消息进行虚假交易信息识别。

2. 如权利要求 1 所述的虚假交易信息识别方法，其特征在于，在获取用户发布信息的信息特征之前，包括以下步骤：

步骤 S1011，获取之前用户发布消息的基本数据；

步骤 S1012，根据获取的之前用户发布消息的基本数据，提取训练数据，确定正负样本；

步骤 S1013，对正负样本中的数据进行特征转换，得到设定数据格式的数据；

步骤 S1014，根据设定数据格式的数据，建立回归模型。

3. 如权利要求 2 所述的虚假交易信息识别方法，其特征在于，步骤 S1013 具体包括：

将正负样本中的每条数据的特征确定为数值型或枚举型两类；

数值型的维度值不变，在数值型数据处于样本中的位置处置该数值型数据的数值；

枚举型的维度值则先计算其 md5 值，然后将 md5 值对 W 取模，得到取模结果；在样本中将处于取模结果位置的数值置 1。

4. 如权利要求 3 所述的虚假交易信息识别方法，其特征在于，步骤 S1014 具体包括：

将步骤 S1013 得到的数据转化为稀疏矩阵；

在模型训练程序程序中输入产生的稀疏矩阵（$x_1, x_2, x_3, x_4, x_5, \cdots\cdots, x_p$），p 为设定数据格式的数据的数据量；得到每一条记录对应的参数（$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \cdots\cdots, \beta_p$）；

建立回归模型，回归模型为：$P(Y=1|x)=\pi(x)=\frac{1}{1+e^{-g(x)}}$；其中 $g(x) = \beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p$。

5. 如权利要求 4 所述的虚假交易信息识别方法，其特征在于，在建立回归模型之后，当接收到用户发布信息时，则步骤 S101 具体为：

步骤 S1015，获取用户发布消息的基本数据；包括提取用户发布消息的基本特征和获取元特征；基本特征与元特征一起作为挖掘的基本数据。

6. 如权利要求 5 所述的虚假交易信息识别方法，其特征在于，在获取用户发布消息的基本数据之后，步骤 S201 具体包括以下步骤：

步骤 S2011，对获取用户发布消息的基本数据进行特征转换，得到设定数据格式的数据；

步骤 S2012，将步骤 S2011 得到的设定数据格式的数据转化为稀疏矩阵的形式，通过回归模型进行虚假消息识别；其中，P>M，则 Y=1，表示用户发布信息为真实交易信息；反之，P ≤ M，则 Y=0，表示用户发布信息为虚假交易信息；M 是预先设定的阈值。

7. 如权利要求 1 或 6 所述的虚假交易信息识别方法，其特征在于，在获取用户发布信息的信息内容之前，包括以下步骤：

步骤 S1021，获取之前用户发布消息的信息内容并进行审核，将通过审核与没通过审核的信息分为两类，作为分类的样本数据；

步骤 S1022，对样本中的信息内容进行分词；

步骤 S1023，通过计算，抽取特征词；

步骤 S1024，计算每类中每篇文档内的每个特征词的特征值；

步骤 S1025，根据获取每类中每篇文档内的每个词的特征值，通过训练得到识别模型。

8. 如权利要求 7 所述的虚假交易信息识别方法，其特征在于，步骤 S1023 具体包括：

对每个词都求 CHI 值；开方检验公式为：$x^2(t,c)=\frac{N(AD-BC)^2}{(A+C)(A+B)(B+D)(C+D)}$ 其中，A：在这个分类下包含这个词的文档数量；B：不在该分类下包含这个词的文档数量；C：在这个分类下不包含这个词的文档数量；D：不在该分类下，且不包含这个词的文档数量；N：表示文章总数；t：表示当前求 CHI 值的词；c：表示分类的类别；x2：表示开放检验 CHI 值；

然后取所有词中 CHI 值最大的 P 个值作为特征词；

步骤 S1024 具体包括：

采用 TFIDF 算法，计算每类中每篇文档内的每个特征词的次数，以及包含这个词的文档数，用 TFIDF 的值作为特征值；其中，将每篇文档转化为：类别 ID\t 特征序号 \t 特征值的格式；TFIDF 公式为：TFIDF=TF×IDF，其中，TF 为某个特征词在这篇文档中出现的频率，IDF 为反文档频率，即总文档树除以包含这个词的文档数。

9. 如权利要求 8 所述的虚假交易信息识别方法，其特征在于，在获取用户发布消息的信息内容之后，步骤 S201 具体包括以下步骤：

步骤 S2021，对用户发布消息的信息内容进行分词；

步骤 S2022，通过计算，抽取特征词；

步骤 S2023，计算用户发布消息的信息内容中的每个词的特征值；

步骤 S2024，根据得到的识别模型，对用户发布消息的信息内容进行虚假交易信息识别。

10. 如权利要求 1、6 或 9 所述的虚假交易信息识别方法，其特征在于，根据用户发布信息的图片信息，对用户发布消息进行虚假交易信息识别，具体包括以下步骤：

步骤 S2031，查询图片库，判断当前图片是否在图片库中出现，如果出现，则进一步判断发帖内容是否相同，以及位置是否相同，如果都不同，则判定包含该图片的用户发布信息是虚假交易信息；否则，则判定包含该图片的用户发布信息是真实交易信息；

或者，判断图片上是否有水印，如果有，则进一步判断图片上的水印是否合法，如果不合法，则判定包含该图片的用户发布信息是虚假交易信息；否则，则判定包含该图片的用户发布信息是真实交易信息。

一种虚假交易信息识别方法

技术领域

[0001] 本发明涉及互联网技术领域，特别是涉及一种虚假交易信息识别方法。

背景技术

[0002] 随着互联网的发展，网上的信息变得越来越泛滥，越来越真假难辨。对于电子商务或分类信息等类型的网站，如果能够为用户提供安全、真实的商品信息，已经成为一项重要而又基本的内容，于是如何识别用户发布信息的真假已经成为了确保信息安全的关键，这也是很多网站都面临的问题。

[0003] 在识别虚假交易信息上，目前的方法主要是通过人工的审核，外加一些技术手段，例如确定黑名单的 IP（Internet Protocol，网络之间互连的协议）地址、确定发布的信息内容或格式不合法、价格区间不合法等将完全确定信息不合法的信息删除。

[0004] 现有策略的缺点是：人工审核太消耗人力、辅助的技术手段只能删除少部分的虚假交易信息，还有大量的虚假交易信息逃脱，可以删除 100% 确定为虚假的信息，但是对有 85% 可能为假的信息无能为力，因为都不能判断信息为假的程度。

发明内容

[0005] 本发明要解决的技术问题是提供一种虚假交易信息识别方法置，用以解决现有技术进行虚假交易信息识别上人工消耗大、虚假交易信息识别率低的问题。

[0006] 为解决上述技术问题，一方面，本发明提供一种虚假交易信息识别方法，包括：

[0007] 步骤 S101，获取用户发布信息的信息特征、信息内容和 / 或图片信息；

[0008] 步骤 S201，根据用户发布信息的信息特征、信息内容和 / 或图片信息，对用户发布消息进行虚假交易信息识别。

[0009] 进一步，在获取用户发布信息的信息特征之前，包括以下步骤：

[0010] 步骤 S1011，获取之前用户发布消息的基本数据；

[0011] 步骤 S1012，根据获取的之前用户发布消息的基本数据，提取训练数据，确定正负样本；

[0012] 步骤 S1013，对正负样本中的数据进行特征转换，得到设定数据格式的数据；

[0013] 步骤 S1014，根据设定数据格式的数据，建立回归模型。

[0014] 进一步，步骤 S1013 具体包括：

[0015] 将正负样本中的每条数据的特征确定为数值型或枚举型两类；

[0016] 数值型的维度值不变，在数值型数据处于样本中的位置处置该数值型数据的数值；

[0017] 枚举型的维度值先计算其 md5 值，然后将 md5 值对 W 取模，得到取模结果；在样本中将处于取模结果位置的数值置 1。

[0018] 进一步，步骤 S1014 具体包括：

[0019] 将步骤 S1013 得到的设定数据格式的数据转化为稀疏矩阵；

[0020] 在模型训练程序中输入产生的稀疏矩阵（$x_1, x_2, x_3, x_4, x_5, \cdots\cdots, x_p$），p 为设定数据格式的数据的数据量；得到每一条记录对应的参数（$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \cdots\cdots, \beta_p$）；

[0021] 建立回归模型，回归模型为：$P(Y=1|x)=\pi(x)=\frac{1}{1+e^{-g(x)}}$；其中 $g(x)=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p$。

[0022] 进一步，在建立回归模型之后，当接收到用户发布信息时，则步骤 S101 具体为：

[0023] 步骤 S1015，获取用户发布消息的基本数据；包括提取用户发布消息的基本特征和获取元特征；基本特征与元特征一起作为挖掘的基本数据。

[0024] 进一步，在获取用户发布消息的基本数据之后，步骤 S201 具体包括以下步骤：

[0025] 步骤 S2011，对获取用户发布消息的基本数据进行特征转换，得到模型可处理的数据格式；

[0026] 步骤 S2012，将步骤 S2011 得到的数据转化为稀疏矩阵的形式，通过回归模型进行虚假消息识别；其中，P>M，则 Y=1，表示用户发布信息为真实交易信息；反之，P ≤ M，则 Y=0，表示用户发布信息为虚假交易信息，M 是预先设定的阈值。

[0027] 进一步，在获取用户发布信息的信息内容之前，包括以下步骤：

[0028] 步骤 S1021，获取之前用户发布消息的信息内容并进行审核，将通过审核与没通过审核的信息分为两类，作为分类的样本数据；

[0029] 步骤 S1022，对样本中的信息内容进行分词；

[0030] 步骤 S1023，通过计算，抽取特征词；

[0031] 步骤 S1024，计算每类中每篇文档内的每个特征词的特征值；

[0032] 步骤 S1025，根据获取每类中每篇文档内的每个词的特征值，通过训练得到识别模型。

[0033] 进一步，步骤 S1023 具体包括：

[0034] 对每个词都求 CHI 值；开方检验公式为：$x^2(t,c)=\frac{N(AD-BC)^2}{(A+C)(A+B)(B+D)(C+D)}$ 其中，A：在这个分类下包含这个词的文档数量；B：不在该分类下包含这个词的文档数量；C：在这个分类下不包含这个词的文档数量；D：不在该分类下，且不包含这个词的文档数量；N：表示文章总数；t：表示当前求 CHI 值的词；c：表示分类的类别；x^2：表示开放检验 CHI 值；

[0035] 然后取所有词中 CHI 值最大的 P 个值作为特征词；

[0036] 步骤 S1024 具体包括：

[0037] 采用 TFIDF 算法或 TFIDF 的变形算法计算特征值，其中 TFIDF 的做法是计算每类中每篇文档内的每个特征词的次数，以及包含这个词的文档数，用 TFIDF 的值作为特征值；其中，将每篇文档转化为：类别 ID\t 特征序号 \t 特征值的格式；TFIDF 公式为：TFIDF=TF×IDF，其中，TF 为某个特征词在这篇文档中出现的频率，IDF 为反文档频率，即总文档树除以包含这个词的文档数。

[0038] 进一步，在获取用户发布消息的信息内容之后，步骤 S201 具体包括以下步骤：

[0039] 步骤 S2021，对用户发布消息的信息内容进行分词；

[0040] 步骤 S2022，通过计算，抽取特征词；

[0041] 步骤 S2023，计算用户发布消息的信息内容中的每个词的特征值；

[0042] 步骤 S2024，根据得到的识别模型，对用户发布消息的信息内容进行虚假交易信息

识别。

[0043] 进一步，根据用户发布信息的图片信息，对用户发布消息进行虚假交易信息识别，具体包括以下步骤：

[0044] 步骤 S2031，查询历史图片库，判断当前图片是否在图片库中出现，如果出现，则进一步判断发帖内容是否相同，以及位置是否相同，如果都不同，则判定包含该图片的用户发布信息是虚假交易信息；否则，则判定包含该图片的用户发布信息是真实交易信息；

[0045] 或者，判断图片上是否有水印，如果有，则进一步判断图片上的水印是否合法，如果不合法，则判定包含该图片的用户发布信息是虚假交易信息；否则，则判定包含该图片的用户发布信息是真实交易信息。

[0046] 本发明有益效果如下：

[0047] 本发明可以大大的减少交易信息的虚假量，提高交易信息的真实性，增加用户体验，同时可以大大减少人力成本。

附图说明

[0048] 图 1 是本发明实施例中一种虚假交易信息识别方法的流程图。

具体实施方式

[0049] 以下结合附图以及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不限定本发明。

[0050] 如图 1 所示，本发明实施例涉及一种虚假交易信息识别方法，包括：

[0051] 步骤 S101，获取用户发布信息的信息特征、信息内容和 / 或图片信息；

[0052] 步骤 S201，根据用户发布信息的信息特征、信息内容和 / 或图片信息，对用户发布消息进行虚假交易信息识别。

[0053] 步骤 S101 中，具体涉及三种情况，第一种是针对用户发布信息的信息特征进行虚假交易信息识别，也就是基于用户特征和行为进行虚假交易信息识别；第二种是针对用户发布信息的信息内容进行虚假交易信息识别，也就是基于帖子文本内容进行虚假交易信息识别；第三种是针对图片信息进行的虚假交易信息识别。

[0054] 首先，描述基于用户发布信息的信息特征进行虚假交易信息识别，在获取用户发布信息的信息特征之前，包括以下步骤：

[0055] 步骤 S1011，获取之前用户发布消息的基本数据。本步骤中，通过拼接数据，分析用户发帖日志，提取出用户发布消息的基本特征；其中，基本特征是指可以直接从之前用户发布消息中提取获得的数据，例如，用户的身份标识（USER ID）、发帖 IP、cookieid、电话号码、时间信息（包括星期、月份、日期）、发帖时长、浏览量、刷新量、发帖城市、发帖类别等特征。然后，根据用户的基本特征，获取元特征；其中，元特征是指在用户的基本特征的基础上，通过统计或计算得到的数据；例如同 IP 发帖数、同 IP 发帖城市数、同用户发帖数、同用户发帖城市数、同 cookie 发帖数、同 cookie 发帖城市数等元特征。基本特征与元特征一起作为挖掘的基本数据。例如，产生这样一条记录 R1（123123，192.168.11.11，DFOKIEBNGIDH1232，18311067654，……）。

[0056] 步骤 S1012，根据获取的之前用户发布消息的基本数据，提取训练数据。本步骤中，

以步骤 S1011 的结果为基础，通过人工审核验证出确定为真实或者虚假的数据，作为正负样本，真实数据为正样本，虚假数据为负样本；例如，将 R1 标记为正样本或者负样本。人工审核过程，可以根据一些共知信息进行人为判断，也可以通过电话等手段进行验证确认。

[0057] 步骤 S1013，对正负样本中的数据进行特征转换，得到设定数据格式的数据。本步骤中，将正负样本中的每条数据的特征确定为数值型或枚举型两类，其中，数值型是指数据本身就是数值；枚举型是指数据本身不是数值，枚举型根据原始维度及取值进行映射得到。例如 USER ID、发帖 IP 等为枚举型的数据，发帖时长、同用户发帖城市数为数值型的数据。数值型的维度值不变；例如，某特征数据为 20，在样本中的位置在第 10 位，则在第 10 个位置上置 20。枚举型的维度值则先计算其 md5（Message Digest Algorithm MD5，消息摘要算法第五版）值，然后将 md5 值对 W（例如 W = 300000）取模，即：用 md5 值除以 300000，得到余数；这样枚举型的值就会落在 1-300000 之间。例如有两个特征：（电话号码、同电话发帖数），对应的值为（18211078765,100），同电话发帖数为数值型的，电话号码为枚举型的，所以同电话发帖数的位置在样本中的位置不变，电话号码计算 md5 值后，对 300000 取模，例如得到 180834，此时这条记录产生的向量为（0,100,0,……,1），其中，在样本中第 180834 的位置上置 1，表示该位置有数值，数值为 1。

[0058] 步骤 S1014，根据设定数据格式的数据，建立回归模型。对回归模型的要求是回归的结果的值域在 [0,1] 之间，或者可以通过计算映射到这个范围内，以下以逻辑回归为例。步骤 S1013 中得到的是一条条的向量，例如（0,0,0,0,0,0,0,0,0,0,12,32,43,.....1,0,0.....1,0,0.....），因为这些向量可能会有 300000 维，表示数据量会相当耗费内存，所以将一条条的向量转化为稀疏矩阵的形式，例如，如果上一条是第一条，则横坐标为 1，相应的稀疏矩阵的格式为：110（相当于纵坐标）12,11132,11243 等。每一条都这么转化之后，在模型训练程序程序中输入是上面产生的稀疏矩阵，输出是每一条记录对应的参数。可简单的理解为如果一条记录是（$x_1, x_2, x_3, x_4, x_5, \cdots\cdots, x_p$），p 为设定数据格式的数据的数据量；通过模型训练程序进行求解，产生（$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \cdots\cdots, \beta_p$）等对应的参数。此时建立回归模型，回归模型可表示为：$P(Y=1|x)=\pi(x)=\frac{1}{1+e^{-g(x)}}$；其中 $g(x) = \beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p$。

[0059] 在建立回归模型之后，当再接收到用户发布信息时，则步骤 S101 具体为：

[0060] 步骤 S1015，获取用户发布消息的基本数据；包括提取用户发布消息的基本特征和获取元特征；基本特征与元特征一起作为挖掘的基本数据。具体内容与步骤 S1011 相同，本步骤不再详细描述。

[0061] 在获取用户发布消息的基本数据之后，步骤 S201 具体包括以下步骤：

[0062] 步骤 S2011，对获取用户发布消息的基本数据进行特征转换，得到设定数据格式的数据。本步骤与步骤 S1013 方法相同，不再详述。

[0063] 步骤 S2012，将步骤 S2011 得到的设定数据格式的数据转化为稀疏矩阵的形式，通过回归模型进行虚假消息识别。本步骤中，得到稀疏矩阵后，根据得到的用户发布消息对应的（$x_1, x_2, x_3, x_4, x_5, \cdots\cdots, x_p$），就可以得到 g(x)，这样就可以求得 P(Y=1|x) 的结果，即 Y=1 的概率；其中，P>M，则 Y=1，表示用户发布信息为真实交易信息；反之，P ≤ M，则 Y=0，表示用户发布信息为虚假交易信息；M 是预先设定的阈值。

[0064] 其次，描述基于用户发布信息的信息内容进行虚假交易信息识别，在获取用户发

布信息的信息内容之前，包括以下步骤：

[0065] 步骤 S1021，获取之前用户发布消息的信息内容，并对上述内容通过审核（人工审核或自动审核），将通过审核与没通过审核的交易信息帖子作为两类，作为分类的样本数据；可以通过专家人工标签及部分准确率极高（高于设置阈值）的算法自动提取正负样本训练集；

[0066] 步骤 S1022，对样本中的信息内容进行分词，可以通过自定义词典的方式优化分词效果。具体的分词方法可以采用现有的分词方法，例如 ICT 分词方法或其它分词方法。

[0067] 步骤 S1023，抽取特征词。本步骤中，过滤掉步骤 S1022 分词中的停词、罕见词、常见词，然后用 CHI（开方检验）等方法选取与类相关度大的特征词。具体选取方法是：对每个词都求 CHI 值，然后取所有词中 CHI 值最大的 1000 个值作为特征词。开方检验公式为：$x^2(t,c)=\frac{N(AD-BC)^2}{(A+C)(A+B)(B+D)(C+D)}$ 其中，A：在这个分类下包含这个词的文档数量；B：不在该分类下包含这个词的文档数量；C：在这个分类下不包含这个词的文档数量；D：不在该分类下，且不包含这个词的文档数量；N：表示文章总数；t：表示当前求 CHI 值的词；c：表示分类的类别；x^2：表示开放检验 CHI 值。

[0068] 步骤 S1024，进行向量化，获取每类中每篇文档内的每个特征词的特征值。本步骤采用 TFIDF 算法，计算每类中每篇文档内的每个特征词的次数，以及包含这个词的文档数，用 TFIDF 的值作为特征值。将每篇文档转化为：类别 ID\t 特征序号 \t 特征值的格式。TFIDF 公式为：TFIDF=TF×IDF，其中，TF 为某个特征词在这篇文档中出现的频率，IDF 为反文档频率，即总文档树除以包含这个词的文档数。

[0069] 步骤 S1025，根据获取每类中每篇文档内的每个词的特征值，通过训练得到识别模型。本步骤中，采用 SVM（support vector machine 支持向量机）、决策树、贝叶斯分类等方式对上述特征值进行训练，步骤 S1024 中已将每篇文档转化为向量的形式，采用分类（Waikato Environment for Knowledge Analysis，怀卡托智能分析环境）程序对这些向量进行训练，可以选择不同的分类方法，例如 SVM、决策树、贝叶斯分类等，产生一个识别模型。SVM、决策树、贝叶斯分类均为现有成熟的训练方法，本步骤不再详细描述。

[0070] 在得到识别模型之后，当再接收到用户发布信息时，则步骤 S101 具体为：

[0071] 步骤 S1026，获取用户发布消息的信息内容，例如，以用户发帖为例，则获取帖子的具体内容。

[0072] 在获取用户发布消息的信息内容之后，步骤 S201 具体包括以下步骤：

[0073] 步骤 S2021，对用户发布消息的信息内容进行分词。

[0074] 步骤 S2022，抽取特征词。本步骤与步骤 S1023 方法相同，因此，不再详细描述。

[0075] 步骤 S2023，进行向量化，获取用户发布消息的信息内容中的每个词的特征值。本步骤与步骤 S1024 方法相同，因此，不再详细描述。

[0076] 步骤 S2024，根据得到的识别模型，对用户发布消息的信息内容进行虚假交易信息识别。本步骤中，通过 SVM、决策树、贝叶斯分类等方式得到的识别模型为现有成熟模型，其识别方法也是现有成熟技术，因此本步骤不再详细描述。

[0077] 最后，描述基于图片信息进行虚假交易信息识别，在获取用户发布信息的图片信息之后，根据用户发布信息的图片信息，对用户发布消息进行虚假交易信息识别（步骤 S201）包括以下步骤：

[0078] 步骤 S2031，查询图片库，判断当前图片是否在图片库中出现，如果出现，则进一步判断发帖内容是否相同，以及位置是否相同，如果都不同，则判定包含该图片的用户发布信息是虚假交易信息；否则，则判定包含该图片的用户发布信息是真实交易信息；或者，判断图片上是否有水印，如果有，则进一步判断图片上的水印是否合法，如果不合法，则判定包含该图片的用户发布信息是虚假交易信息；否则，则判定包含该图片的用户发布信息是真实交易信息。

[0079] 另外，上述三种策略也可以进行组合，结合一起进行判断，例如，两种情况组合，或三种情况组合；当上述三种情况中有任意一种或两种情况判定用户发布信息是虚假交易信息，则判定用户发布信息是虚假交易信息。

[0080] 由上述实施例可以看出，本发明可以大大的减少交易信息的虚假量，提高交易信息的真实性，增加用户体验，同时可以大大减少人力成本。

[0081] 尽管为示例目的，已经公开了本发明的优选实施例，本领域的技术人员将意识到各种改进、增加和取代也是可能的，因此，本发明的范围应当不限于上述实施例。

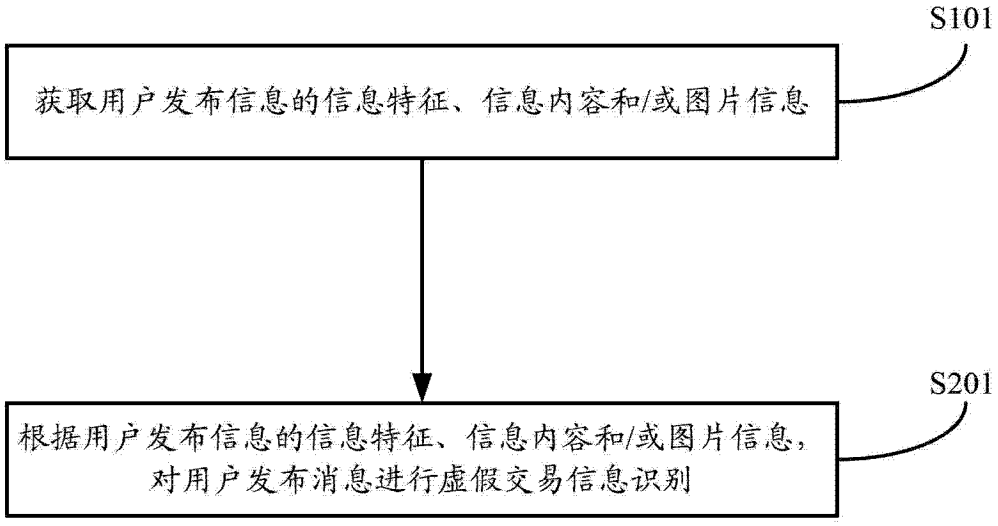

S101
获取用户发布信息的信息特征、信息内容和/或图片信息
S201
根据用户发布信息的信息特征、信息内容和/或图片信息，对用户发布消息进行虚假交易信息识别

图 1