

基于云计算的食品有毒有害物质检验检测大数据的风险分析算法及其应用

王雅洁, 杨冰, 代姣, 何锦林, 陈恺, 罗艳, 谭红, 陶光灿* (贵州省分析测试研究院, 贵州贵阳 550002)

摘要 基于云计算, 结合食品安全检验检测的完备性与最小性原理, 将影响食品安全的多维因素降维成平均含量(AVE)、限量标准(STA)、超限率(OUT)、超限程度(OD)和最大值(MAX)5个因素, 并建立食品有毒有害物质检验检测大数据的风险分析算法。利用云计算技术实现对地理上分布广泛、动态、复杂性高的海量数据进行存储, 并运用云计算的MapReduce计算框架进行智能的并行数据处理及计算, 最后得到风险分析结果。通过对在基于Web端的实验室管理系统采集的1 000 000条检验检测数据结果进行风险分析, 得出食品安全指数IFS远小于1, 表明消费者人群的食品安全状态良好。

关键词 云计算; 食品有毒有害物质检测; 大数据; 风险分析算法

中图分类号 S126; R155 **文献标识码** A **文章编号** 0517-6611(2017)21-0216-05

Risk Analysis Algorithm and Its Application of Poisonous and Harmful Substance in Food Testing Big Data Based on Cloud Computing
WANG Ya-jie, YANG Bing, DAI Jiao, TAO Guang-can* et al (Guizhou Academy of Testing and Analysis, Guiyang, Guizhou 550002)

Abstract Based on cloud calculation, combined with the completeness and minimum principle in food safety inspection and testing, multiple dimensional factors that affected the food safety were reduced into five factors: average content (AVE), limit standard (STA), overload rate (OUT), out of limit degree (OD), and the maximum value (MAX), and the poisonous and harmful substance risk analysis algorithm in food safety inspection and testing big data was established. The paper made use of cloud computing platform to achieve the data storage of massive extensive geographical distribution, dynamic, high complexity data, and applies MapReduce computational framework of cloud computing for intelligent parallel data processing and computing. Finally, we got the required risk analysis results. Through the risk analysis of collected 1 000 000 testing data results from the laboratory management information system based on web side, it was found that the food safety index was greater less than 1, which indicated that the food safety state was in good condition in consumer population.

Key words Cloud computing; Food testing of poisonous and harmful substance; Big data; Risk analysis algorithm

近年来,随着互联网^[1]的高速发展,云计算技术^[2]已在金融^[3-5]、汽车^[6-7]、电力^[8-9]等行业得到一定的应用。李聪^[10]提出的食品安全状态评价指标体系曾在食品安全风险分析中有所应用,但由于风险分析的数据量过小,分析结果不具有说服力,且并未实现在云技术下的应用。目前,在食品行业“互联网+食品检验检测”模式驱动下,食品安全检验检测数据量呈指数级增长,大数据已经形成。大量的食品检测数据无疑能为人们带来广阔的信息量,但需要从海量食品安全检验检测数据中发现对监管部门、企业及检测机构有用知识的难度随之增加。食品安全风险分析是食品安全领域的重难点,基于云计算的食品有毒有害物质的风险评估算法能够有效解决海量数据潜在价值的挖掘及利用。笔者通过采集互联网端实验室信息管理系统(Laboratory management information system, LIMS)汇聚的全国各检验检测机构的食品检验检测大数据,并充分运用云计算平台高可用性 & 高度虚拟化等特征^[11],动态调度和分配资源,以满足精准风险分析及高效数据挖掘的需求。笔者提出一种基于云计算的食品有毒有害物质检验检测大数据的风险分析算法,将数学算法应用在食品安全检验检测中。该模型是一种使食品安全检验检测结果反映数据间呈现出多变量间关系的一种算法应用技术,该算法已在贵州省食品安全云平台实际部署并上线运行,取得了良好的效果。该文是“互联网+食品安全检验检测”模式的实践应用,实现了一种以海量数据信息为中心

的风险交流方法。

1 云计算的定义及云计算环境下的并行计算模型

1.1 云计算的定义 云计算^[12]是一种实现海量数据的分布式计算模式,通过聚合海量分布在不同时间、不同地点、不同食品安全检验检测实验室的食品检测结果的资源,提供海量食品数据的存储与计算能力。云计算平台通过虚拟化、动态资源调配等技术向食品行业的监管部门、企业、检测机构、媒体、大众提供数据计算的按需服务,避免资源浪费,提高云服务器的利用率以及应用性能。云计算使用大量廉价互连在互联网上的计算机按需进行任务的处理,为食品安全检验检测结果大数据提供了所需的存储资源、计算资源等。所有用户无需了解云计算基础设施的管理。与此同时,云计算提供动态负载均衡与横向伸缩的能力,随着数据量的增大,需要更多的数据节点来进行计算。云计算支持运行时向数据中心按需新增节点,并自动将一些负载按需转移到新的节点上,进行运算,并保持节点间负载的平衡,提高业务的承载能力^[13]。

此外,利用虚拟化技术原理,云计算能实现分布式计算、信息资源的集中管理和智能调配。云计算主要有3种主流的商业模式,分别为平台即服务(PaaS)、基础架构即服务(IaaS)和软件即服务(SaaS)。云计算逻辑图如图1所示。

在云计算环境中,所有数据资源分为多个数据中心的模式。一个数据中心有成千上万的数据节点,这些数据节点间通过互联网进行高效连接,可为用户提供存储及计算资源。随着云计算的快速发展,能够利用虚拟化实现廉价大规模的计算平台,并将存储资源、网络资源、计算资源、应用程序等作为虚拟化的实体。虚拟化技术可被运用于有效利用和组

基金项目 “十二五”科技支撑计划项目(2015BAK36B04)。

作者简介 王雅洁(1990—),女,贵州贵阳人,助理研究员,硕士,从事云计算与大数据挖掘研究。*通讯作者,副研究员,博士,从事食品安全风险分析工作。

收稿日期 2017-05-12

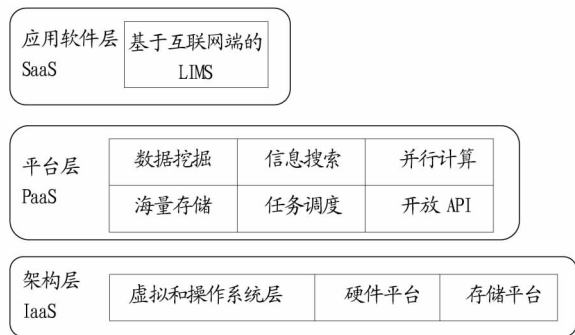


图1 云计算逻辑

Fig.1 The logic of cloud computing

织空闲的计算平台,对闲置的计算资源进行抽象,使计算资源形成相互间独立的虚拟服务器实例,从而独立地完成被分配的数据处理和计算,以此实现底层硬件的虚拟化。

1.2 云计算环境下的并行计算模型 MapReduce 是 Google 实验室提出的一个云计算环境下处理大规模数据的分布式并行计算模型框架^[14-15]。MapReduce 作业处理分为 Map 步骤和 Reduce 步骤,任务由大量独立并行的 Map 任务和 Reduce 任务组成,每个步骤可用一个函数来表示,即 Map 函数和 Reduce 函数,每个函数都以一个键值对 <key, value> 的形式作为输入和输出。Map 步骤将大规模数据处理作业分成多个可独立运行的小任务,与此同时,Map 函数会输入一个键值对 <key, value>,并通过虚拟化技术,向计算集群中的节点计算机分配多个子任务,进行分布式处理。Reduce 函数则对 key 值相等的 value 进行处理,生成键值对集合,Reduce 步骤将若干个任务的处理结果进行汇总,最后输出分析结果。在开发 MapReduce 并行计算模型的过程中,首先对通过互联网应用程序采集到的海量数据进行读取,将大规模需要处理和计算的数据拆分成若干数据块,之后用户编写相应的 MapReduce 执行程序,通过分布式计算机集群运行,对大规模的数据集合进行分解,通过形成一系列数据片段的方式得到若干键值对,这样的过程通过定义 Map 及 Reduce 的接口即可实现。MapReduce 编程模型下产生的结果最终写入 HDFS 中。

基于 MapReduce,云计算可实现大规模数据的高效处理及分析。大规模数据在使用 MapReduce 计算框架进行并行化计算的时候,需保持结构一致。对于海量数据的挖掘,通常跟计算难度大的复杂算法有关,比如近似求解、数据降维、程序迭代等。根据集体行业不同,往往需要根据行业数据定制化数据挖掘算法。因此,基于云计算和具体行业的海量数据挖掘技术引起了学术界和各行业的高度关注。目前,随着数据量的增长,数据信息量通常能够达到 GB 或 TB 级别。云计算平台采用 HDFS 对大规模数据进行支持,能在大规模分布式的廉价计算集群节点中进行扩展且提供高数据的带宽。

2 食品有毒有害物质检验检测大数据的风险分析

基于云计算的大数据挖掘分析技术目前已被运用于互联网行业的软件架构与开发,软件设计、功能,能够充分表达出互联网大数据特征以及业务逻辑。当应用于除互联网外

的具体行业时,首先需要对特定行业的大数据特征进行分析,然后再根据业务逻辑进行定制开发。为了满足对食品安全检验检测大数据深入分析、挖掘的需求,针对食品检验检测行业的业务需求与大数据特征,通过部署在互联网端的食品安全 LIMS,采集基于云平台的跨地区、跨地点、不同实验室的“互联网+食品安全检验检测”实验室大数据,通过数据 ETL,并结合云计算技术的最新进展与实际行业部署经验,在 PaaS 层(图1)设计了一种食品有毒有害物质检验检测大数据的风险分析模型,通过云计算的 MapReduce 计算框架进行分布式计算及分析,最终得到分析结果,该结果可通过 SaaS 层(图1)基于互联网端的软件系统进行可视化展示。

2.1 数据的备份 为保证食品安全检验检测数据的高可靠性应用,采用云计算的分布式存储方式来存放数据,针对通过互联网端实验室管理系统采集的不同时间、不同地点、不同实验室的食品安全检验检测大数据,将同一份数据存储为多个副本,具体操作时将检验检测结果大数据切分成不同的数据块大小,而不是基本的复制,之后再基于不同的配置信息将若干个数据块存储到相应的数据节点上。

2.2 计算向存储迁移 研究基于云计算的食品有毒有害物质检验检测大数据的风险分析是研究一种基于食品行业的数据挖掘算法。设计的数据挖掘算法是基于海量检测结果的,在分布式系统中,网络带宽资源较为稀缺,各节点的计算能力则较为富裕。在具体任务调度的运行过程中,云计算平台将存储节点与计算节点放在一起,并且尽量在存放对应输入数据块的分布式节点上分配及执行具体任务,这种机制使并行任务在本地计算机节点上读取输入数据,可在一定程度上减少网络数据流^[16]。

2.3 数据挖掘 数据挖掘也称为知识发现过程,就是从采集的大规模数据中,通过挖掘算法,发现潜在可被理解的、有价值的信息的过程。对于食品有毒有害物质的风险分析是数据挖掘的一种具体实现,最终的目的是从海量检验检测数据中提取出可理解的知识,挖掘出潜在价值。因此,希望数据规模的越大越好,这样数据挖掘出结果更加精准,即风险分析更加有说服力。这样高要求的风险分析对于开发环境和应用环境有较高的要求,基于云计算的风险分析方式较为满足。在云计算平台中,大规模的数据被存储在数据中心中,云计算根据风险分析算法的具体应用需求动态分配资源进行计算处理,并采用自带容错机制确保数据挖掘算法的可靠性及可扩展性^[17]。

① 基于云计算模式可使风险分析分布式并行的执行,有效避免了资源浪费,节约运算时间。该模式可运用在不同规模的行业组织中,为食品企业、监管部门、检测机构带来新型低成本的计算环境,减轻行业组织对成本大型昂贵的高性能机的依赖性。② 基于云计算的风险分析算法开发方便,由于底层被屏蔽掉,行业组织不再需考虑底层数据的划分、数据节点的分配以及云计算任务调度等。③ 基于云计算模式能很大程度上提升海量数据的处理能力,可按需智能化的增加结点,提高容错性。④ 基于云计算的风险分析保证了数据

挖掘技术的共享,降低了行业数据挖掘应用门槛,满足了大规模食品检验检测数据的价值提取。

2.3.1 系统架构。基于以上思想,设计了基于云计算技术的数据挖掘系统,基于食品有毒有害物质的风险分析算法即为数据挖掘算法库中的一种,其整体结构如图2所示。

在设计的数据挖掘系统中,节点被分为主控节点(TotalCtrlNode)及工作节点(WorkerNode)。在整个系统架构中只有一个TotalCtrlNode,由HDFS中的NameNode、SecondNameNode、数据仓库、MapReduce计算模型下的JobTracker、以及数据挖掘算法库组成,设计的风险分析算法即为数据挖掘算法库中的一种,随着业务系统的需要,此后可在数据挖掘算法库中设计更多的基于食品安全行业的风险分析算法。其中,NameNode是HDFS的主服务器,用于管理及存储文件系统的元数据,执行HDFS打开、关闭、重命名等命名空间操作,并将文件块分成若干个,到WorkerNode中的DataNode进行映射。在设计的数据挖掘系统中,可按需部署若干个WorkerNode节点,它由HDFS中的DataNode以及MapReduce计算模型下的TaskTracker组成,分别负责数据的存储和计算工作。其中将实际的大规模数据分成数据块存放在DataNode中,DataNode还负责按照NameNode的命令,处理实际的读写请求,并执行数据块的创建、复制、删除等工作。TaskTracker则利用开发的风险分析算法,处理若干数据块的实际数据挖掘工作。在实际系统运行过程中,首先从架构在云端、基于互联网的应用系统LIMS中采集的不同地点、不同时间、不同实验室的海量检测数据,并存入数据仓库中,此时基于云计算模式下HDFS的NameNode会自动将数据文件分成若干块存储到WorkerNode中的各个DataNode中。SeondNameNode和NameNode部署及运行在不同的机器上,其功能主要是辅助NameNode处理映像文件和事务日志。

数据挖掘算法库负责存储数据挖掘所需的算法,开发了基于食品有毒有害物质检验检测大数据的风险分析算法,利用云计算并行执行的特点,算法使用MapReduce编程模型重新设计实现。

基于食品有毒有害物质检验检测大数据的风险分析算法放入TotalCtrlNode的算法库中,TotalCtrlNode节点中MapReduce计算模型下的JobTracker会进行各计算节点之间的工作进程的调度与协调,按需将数据挖掘任务跨WorkerNode节点进行分发、监控以及失败任务的重新执行,TaskTracker执行时将会按需调用这些任务,或根据具体业务需要结合本地数据库执行TotalCtrlNode分配的任务,并将计算结果及相应状态信息向TotalCtrlNode进行反馈。JobTracker和TaskTracker采用Master/Slave模式进行工作,Master和Slave通过相互发送命令实现交互。这样的结构机制具有高容错性,由于系统中可按需进行WorkerNode的扩展利用,当其中某个WorkerNode出错时,只需将此WorkerNode上正在执行或计算的任务向其他WorkerNode上进行迁移,并重新执行坏掉的WorkerNode上的计算任务,与此同时,向TotalCtrlNode发送迁移信息,TotalCtrlNode将重新对其他的WorkerNode节点进行任务再分配,在此过程中坏掉的WorkerNode将被屏蔽在系统外等待修复。

设计的基于云计算技术的数据挖掘系统可按功能上划分成2个部件:存储部件和计算部件。其中,系统存储部件由HDFS中的NameNode、SecondNameNode、DataNode以及基于互联网端的LIMS数据的数据仓库组成,而计算部件则由MapReduce计算模型下的JobTracker、TaskTracker以及本文设计的数据挖掘算法库组成。该基于云计算技术的数据挖掘系统是Hadoop作为底层大数据解决方案,使用部署在云端的、基于Linux操作系统的计算机集群作为硬件设备。

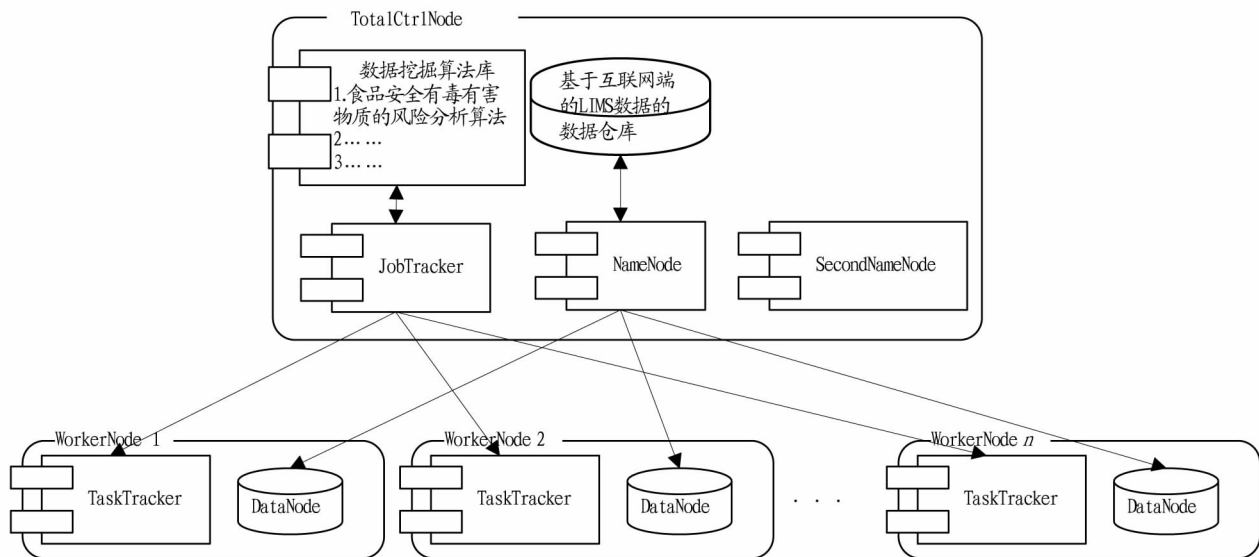


图2 基于云计算的数据挖掘系统架构

Fig.2 Architecture of data mining system based on cloud computing

2.3.2 数据挖掘算法层-风险分析算法设计。基于食品安全检验检测结构化的完备性与最小性的原理^[3],在基于云计

算的数据挖掘系统中设计了基于云计算的食品有毒有害物质检验检测大数据的风险分析算法。由于影响食品安全的

因素繁多,涉及的内容较广,因此对检验检测结果风险分析需要从全局视角出发,从不同层次来考虑食品安全的总体状态。风险分析的原理主要是根据风险分析与食品安全状态影响的重要程度、风险分析与食品安全状态变化的协调性、风险分析反映食品安全状态变化的可靠灵敏性和风险分析刻画食品安全状态变化的代表性来进行评估的^[18]。完备性就是能在食品安全状态变化中找到相应的结构化量度,并能满足食品安全检验检测数据分析结构化的需要;最小性则是对任意结构集 $Q \supset X$, 均有 $X \supset I$, 使得结构集在精减到最小限额结构化前提下依然能获得几乎与其它结构集同样的信息来满足食品安全数据分析结构化的需要^[3]。

2.3.2.1 以阈值法(Threshold approach)应用的技术原理。食品有毒有害物质检验检测大数据结构化的计算,其理论依据是人体与试验动物存在着合理可比的阈值量值^[19]。但是,人的敏感性与动物相比更高,遗传差异性较大,并且膳食习惯更为不同。鉴于此,可采用 ADI^[20]、PTWI^[21] 或 RDI^[22] 等数据进行。

2.3.2.2 食品有毒有害物质检验检测大数据集的风险分析。风险分析涉及到数据集中的具体检测指标,例如六六六在蔬菜中的含量、甲醛在啤酒中的含量等,检测结果大数据是整个风险分析的最底层和最基础的数据。确定大数据集的风险分析非常简单方便,主要根据国内外所制定的有关各类食品的检验标准或法规,针对有毒有害指标,通过基于 Web 端的实验室检测管系统理(LIMS),利用互联网采集跨地区、跨实验室的相关检验检测大数据,并设计实现以下算法,利用互联网采集的原数据和云计算技术,计算相关参数,从而进一步进行风险分析。

(1)影响食品安全的多维因素降维。结合食品安全检验检测结构化的完备性与最小性原理,首先将影响食品安全的多维因素降维成以下几个因素^[3],并进行相关计算,这 5 个因素即为风险评价指标。

①平均含量(AVE)。平均含量 AVE_{item} 指在某类有毒有害物质在某种特定食品中的的平均值,也是具体某个有毒有害物质检测项目结果都数据平均值。其计算公式为:

$$AVE_{item} = \frac{\sum_{i=1}^n C_i}{n} \quad (1)$$

式中, AVE_{item} 为检测指标数据的平均值,由某一有毒有害物质的检验检测指标值计算得到, C_i 表示第 i 个样品中该检测结果的含量, n 为样本总个数。

②限量标准(STA)。针对每一个检测结果指标,按照国际或国家标准,标准中会规定该指标的限量标准(STA_{item})。由于在食品安全领域,有毒有害物质、致病微生物等有害物的风险分析通常是人们所关注的,因此限量标准(STA_{item})在食品安全检验检测的风险分析中的都应该为上限标准。对于标准中规定为“没有”或“未检出”等非数字的情况,可以将其视为“0”处理。

③超限率(OUT)。超限率是在整个的数据集中,该检测项目结果数据值超过限量标准(STA_{item})的样品个数占整个

结果数据集的比例,计算公式如下:

$$OUT_{item} = \frac{M}{N} \times 100\% \quad (2)$$

式中, OUT_{item} 为超限率, M 为超出限量标准的数据个数, N 为整个数据集中的数据个数。

④超限程度(OD)。超限程度(OD_{item})代表在整个的监测结构集中,所有超限数据超出限量标准(STA_{item})的偏离程度。

$$OD_{item} = \frac{\sqrt{\sum_{j=1,2,\dots,m} (C_j - STA_{item})^2}}{M} \quad (3)$$

式中, OD_{item} 代表超限程度, C_j 为在该指标结构化上超限的某一具体数据值, STA_{item} 指该指标的限量标准, M 代表超出限量标准的数据个数。

⑤最大值(MAX)。计算在整个数据集中,该检测指标项检测结果的最大值(MAX_{item})。

(2)食品种类指标。设计了针对具体某个食品中具体检测项目的 5 个评价指标,解决了具体食品中具体项目的风险分析问题。由于各种食品的构成非常复杂,其中可能含有多种致病微生物和多种有毒有害物质,要实现对某一类食品检验检测风险分析,需要在指标基础上建立食品种类指标。食品种类的指标风险分析主要有以下 2 个方面。

①食品的合格率(σ)。针对具体某个食品,比如肉制品,根据其各个检测项目的具体结构化的数据结果,可根据与国家的肉制品评价标准比对,评价出该肉制品样品是否达到国家标准,即是否合格。不同种类的食品合格性评价的标准和方法不同,比如在某些食品中,只要有任何一个检测项目的结构化结果超出其限量标准就可判定为这整个食品不合格。在有些食品中,则需要综合考虑多个检测项目后,再进行评价。无论哪种方式,所有样本都可以作出合格与否的评价,因此可以得到整个云计算数据集中,该食品的合格率(σ_{food}),计算公式为:

$$\sigma_{food} = \frac{L}{N} \times 100\% \quad (4)$$

式中, σ_{food} 为食品合格率, L 表示在整个数据集达到国家标准的样品量,即合格的样品量, N 表示数据集中的样品总量。 $(1 - \sigma_{food})$ 则为该食品的不合格率。

②食品的不安全度(h)。食品的不安全度(h)表示具体某食品整体的不安全程度,是一个无量纲的数值。由于食品中往往有许多微生物或化学物质等危害安全的物质,所以要用一个结构化来表示该类食品的不安全程度,肯定要对食品中有害物质的含量进行无量纲化处理,同时还需要知道该类有害物质的危害系数^[23],食品的不安全度(h_{food})的计算公式为:

$$h_{food} = \sum_{AVE_{item} > STA_{item}} \left(\frac{AVE_{item} - STA_{item}}{STA_{item}} \times F_{item} \right) \quad (5)$$

式中, h_{food} 为某类食品不安全度, F_{item} 代表该有毒有害物质或致病微生物项目的危害系数,通常情况下, $F_{item} \leq 1$, AVE_{item} 为检测指标数据的平均值, STA_{item} 为该指标的限量标准;公式

(5)表示在该食品所有危害物结构化中,挑出其平均含量 AVE_{item} 超出限量标准 STA_{item} 的结构化,将所有超限危害物结构化进行无量纲化处理,即 $\frac{AVE_{item} - STA_{item}}{STA_{item}}$, 然后与该项目的危害系数 F_{item} 相乘,最后累加求和得到该食品的不安全度 (h_{food})。

(3)食品安全检验检测整体状态数据分析。食品安全检验检测整体状态数据分析主要是基于暴露评估方法的食品安全指数 (IFS), IFS 表示食品中的危害物时消费者健康是否存在危害以及危害的程度,即 IFS 能对食品安全整体状态进行评估。

$$IFS = \frac{\sum_{c=c_1, \dots, c_n} IFS_c}{n} \quad (6)$$

式中, IFS_c 表示食品中某种危害物质 C 对消费者健康影响的单项食品安全指数, n 代表样品总个数。

由于食品中各种危害物的毒害作用与其进入人体的绝对量有关,运用阈值法 (Threshold approach) 应用的技术原理,在设计食品有毒有害物质风险分析的评价时,认为评价食品安全整体状态以人体通过饮食途径而对危害物质的实际摄入量与其安全摄入量比较更为科学合理。通常对长期动物试验资料的安全系数为 $100^{[19]}$, 但不同的国家采用的是不同的标准,一般采用 $ADI^{[20]}$ 、 $PTWI^{[21]}$ 、或 $RDI^{[22]}$ 数据。在这样一种理论思想指导下,首先导出可以用来评价食品中某种危害物 C 对消费者健康影响的食品安全指数 IFS_c , 结构化的计算公式为:

$$IFS_c = \frac{EDI_c \times f}{SI_c \times bw} \quad (7)$$

式中, SI_c 为食品安全摄入量,可根据不同的危害物可采用 $ADI^{[20]}$ 、 $PTWI^{[21]}$ 、或 $RDI^{[22]}$ 标准进行取值; bw 为平均体重 (kg), 缺省值为 60; f 为校正因子,如果安全摄入量采用 ADI 、 $PTDI$ 等日摄入量数据, f 取 1; 如果安全摄入量采用 $PTWI$ 等周摄入量数据, f 取 7; EDI_c 为化学物质 C 的实际日摄入量估算值, EDI_c 可按照公式 (8) 计算:

$$EDI_c = \sum (R_i \times F_i) \quad (8)$$

式中, R_i 为食品 i 中化学物质 C 的残留水平,即化学物质 C 的检测结果,单位为 mg/kg ; F_i 是食品 i 的估计日消费量,单位是 $g/(人 \cdot d)$; 式 (7) 可用来计算危害物 C 的食品安全指数 IFS_c , 根据计算结果可以得出该危害物对食品安全的影响程度。当 IFS_c 远小于 1, 危害物对食品安全没有影响; 当 $IFS_c \leq 1$, 危害物 C 对食品安全影响的风险是可以接受的; 当

$IFS_c > 1$, 危害物对食品安全影响的风险超过了可接受的限度。考虑了所研究的消费者人群的饮食习惯及各种食品和化学物质的残留情况下, IFS_c 值就具备了可加和性:

$$\overline{IFS} = \frac{\sum_{c=c_1, \dots, c_n} IFS_c}{n} \quad (9)$$

在这种情况下,当 \overline{IFS} 远小于 1, 证明食品安全状态很好; 当 $\overline{IFS} \leq 1$, 食品安全状态较好, 可接受; $\overline{IFS} > 1$, 消费者人群的食品状态不可接受。无论以上哪一种情况, 对于任何一个 IFS_c 值超过 1, 都说明消费者人群对化学物质 C 的暴露已超过了可接受的程度, 应该进入风险管理程序。

3 应用实例

基于云计算技术的数据挖掘同传统数据挖掘的基本过程一致。系统由 6 台物理计算机组成, 其中 1 台为 TotalCtrlNode, 另 5 台为 WorkerNode。通过从基于互联网的应用系统 LIMS 中, 采集不同时间、不同地点、不同实验室 9 月份某贵州本土品牌肉制品的检测数据 1 000 000 条, 对食品污染物铅 (Pb)、汞 (Hg)、砷 (As) 的检测结果进行风险分析, 采集结果汇集了遍布贵州省 9 个实验室的检测数据。进行数据挖掘与分析工作, 以期进行食品有毒有害物质检验检测大数据的风险分析及评估。采集到的数据先被存放于 TotalCtrlNode 中的数据仓库中, 由 HDFS 中的 NameNode 进行数据分块, 并将若干个数据块分配到 5 台 WorkerNode 节点中的 DataNode 进行存储。基于在 TotalCtrlNode 节点中的数据挖掘算法层设计的基于食品有毒有害物质的检验检测风险分析算法, 由 MapReduce 云计算模型的 JobTracker 进行不同数据块的分析计算任务分配, 由 TaskTracker 进行分析计算任务的执行, 所得结果存储在若干个 DataNode 当中。

根据国家标准 GB2762, 污染物铅 (Pb)、汞 (Hg)、砷 (As) 的限量标准 STA 分别为 0.50、0.05 和 0.50 $mg/kg^{[24]}$ 。在 1 000 000 组检测数据中, 铅 (Pb) 检测结果不合格的有 3 组, 汞 (Hg) 不合格的有 2 组, 砷 (As) 不合格的有 2 组。不合格的指标检测结果分别如下: 铅 (Pb): 0.60、0.65、0.66 mg/kg ; 汞 (Hg): 0.06、0.068 mg/kg ; 砷 (As): 0.60、0.611 mg/kg 。

应用实例中, 采用 $ADI^{[20]}$ 摄入量数据, 铅 (Pb)、汞 (Hg)、砷 (As) 的分别为 0.003 56、0.000 1 和 0.002 15 mg/kg 。该肉制品日均估计消费量为 100 g, 即 0.1 kg。

根据设计的基于云计算的食品有毒有害物质的大数据的风险分析算法, 对采集的肉制品检测数据进行了云计算, 得到表 1。

表 1 肉制品有毒有害物质检测的大数据风险分析结果

Table 1 Big data risk analysis results of poisonous and harmful substance in meat products' testing

重金属 Heavy metals	AVE_{item} mg/kg	STA_{item} mg/kg	OUT_{item} %	OOD_{item}	MAX_{item} mg/kg	σ_{food} %	$1 - \sigma_{food}$ %	h_{food}	\overline{IFS}
铅 Pb	0.069	0.50	0.3	0.080	0.660	99.7	0.3	—	0.032 4
汞 Hg	0.022	0.05	0.2	0.010	0.068	99.8	0.2	—	0.368 3
砷 As	0.008	0.50	0.2	0.007	0.611	99.8	0.2	—	0.006 0

地的建设,积极推进天然草地改良和人工种草,提高其产草量和质量,并落实好草地禁牧、休牧、轮牧与草畜平衡及草地保护等基本制度。要发动广大养羊户充分利用冬闲田、农闲田地种植优良牧草作冬春饲料补充。同时,充分利用丰富的农作物秸秆制作青贮、微贮饲料等来喂羊,变废为宝,满足山羊饲料供给。

4.5 加强疫病防治工作 根据全县羊疫病发生、流行的规律,制定科学合理的山羊免疫程序,推行计划免疫。要进行分类安排、指导,在做好强制免疫的基础上,按当地的实情做好其他疫病的免疫工作。疫病防治中要重点抓好规模养殖场、养殖大户、养殖小区的防疫、免疫,要做到应免尽免,不留空档。同时,要强化引种管理,做到规范化引种,进一步建立健全山羊免疫效果监测及流行疫病监测、检测、净化机制,将养羊的疫病风险降到最低。

4.6 强化品牌营销和深加工建设 积极扶持和组织相关单位、企业(组织)开展山羊产品的品牌建设,树立品牌意识。要积极开展产品“三品一标”认证,制定相应产地保护措施、

产品质量标准,将永胜县山羊产品打造为优质品牌农产品,发挥品牌效益。同时要注重市场主体的培育和销售渠道的建设,完善配套服务体系,打造流通绿色通道。另外,还要积极扶持龙头企业(组织)开展产品营销及深加工,向产品的深度和广度进军,延伸产业链,形成产、加、销、农、工、贸一体化发展经营^[7]。

参考文献

- [1] 云南省畜禽遗传资源委员会. 云南省畜禽遗传资源志[M]. 昆明:云南科技出版社,2015:219-222.
- [2] 杨和平. 牛羊生产[M]. 北京:中国农业出版社,2001.
- [3] 戎俊友. 武定县黑山羊产业发展的现状及对策[J]. 云南畜牧兽医,2013(5):22-24.
- [4] 徐成敏,匡国良. 双柏县云岭山羊产业化发展研究[J]. 现代畜牧科技,2016(1):148-149.
- [5] 尚克勤. 转变经营方式是实现我国肉山羊业可持续发展的有效途径[C]//畜牧兽医事业的发展 and 人类健康——21世纪畜牧兽医生产和科学技术论坛论文集. 北京:中国畜牧兽医学会,1999.
- [6] 马友记,李发弟. 中国养羊业现状与发展趋势分析[J]. 中国畜牧杂志,2011,47(14):16-20.
- [7] 农业部. 全国草食畜牧业发展规划(2016~2020年)[J]. 北方牧业,2016(24):35.
- [7] 李石清. 基于 Hadoop 的汽车后市场商品数据平台设计[D]. 广州:华南理工大学,2013.
- [8] 张林鎰,辛献杰,崔冰,等. 面向汽车产品设计的虚拟现实服务平台研究[J]. 系统仿真学报,2014,26(10):2407-2411.
- [9] 吴凯峰,刘万涛,李彦虎,等. 基于云计算的电力大数据分析技术与应用[J]. 中国电力,2015,48(2):111-116.
- [10] 李聪. 食品安全监测与预警系统[M]. 北京:化学工业出版社,2006.
- [11] WANG Q L, VARELA C A. Impact of cloud computing virtualization strategies on workloads' performance[C]//IEEE International Conference on Utility & Cloud Computing. Melbourne:IEEE Computer Society, 2011:130-137.
- [12] BUYYA R, YEO C S, VENUGOPAL S, et al. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility[J]. Future generation computer systems, 2009, 25(6):599-616.
- [13] ARMBRUST M, FOX A, JOSEPH A D, et al. Above the clouds: A Berkeley view of cloud computing[J]. Eecs department university of california berkeley, 2009, 53(4):50-58.
- [14] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters. [J]. Communications of the acm, 2008, 51(1):107-113.
- [15] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters. [J]. Communications of the acm, 2008, 51(1):107-113.
- [16] 应毅,任凯,刘正涛. 基于云计算技术的数据挖掘[J]. 微电子学与计算机, 2013, 30(2):161-164.
- [17] 贺瑶,王文庆,薛飞. 基于云计算的海量数据挖掘研究[J]. 计算机技术与发展, 2013(2):69-72.
- [18] 黄晓娟,刘北林. 食品安全风险预警指标体系设计研究[J]. 哈尔滨商业大学学报(自然科学版), 2008, 24(5):621-623.
- [19] 蒋定国. 铅、镉污染与膳食暴露评估的实证研究[D]. 北京:中国疾病预防控制中心, 2008.
- [20] LU F C. A review of the acceptable daily intakes of pesticides assessed by WHO[J]. Regulatory toxicology & pharmacology rtp, 1995, 21(3):352-364.
- [21] HERRMAN J L, YOUNES M. Background to the ADI/TDI/PTWI[J]. Regulatory toxicology & pharmacology, 1999, 30(2):109-113.
- [22] HEGSTED D M. Recommended dietary intakes of elderly subjects. [J]. American journal of clinical nutrition, 1989, 50(S5):1231-1235.
- [23] 刘为军. 中国食品安全控制研究[D]. 杨凌:西北农林科技大学, 2006.
- [24] 田洪芸,任雪梅,冯伟,等. GB2762-2012《食品安全国家标准食品中污染物限量》的解读[J]. 中国科技博览, 2013(25):295-296.

(上接第220页)

在计算食品安全程度(h_{food})时,由于得到的 $AVE_{\text{item}} < STA_{\text{item}}$,因此没有食品安全程度。由于得到的关于污染物铅(Pb)、汞(Hg)、砷(As)的IFS远小于1,可以得出消费者人群的食品状态很好的结论。

4 小结

通过采集基于互联网端应用系统 LIMS 的食品安全检验检测数据,在云平台 PaaS 层设计了一种基于云技术的食品有毒有害物质检验检测大数据的风险分析算法,该算法是在云技术基础上设计实现基于 MapReduce 的编程模型,与传统食品安全风险分析相比,采用云计算模式使得成本更加低廉、计算更加快速、节点增加容易,且由于云计算的高容错性,可避免因节点损坏造成的数据丢失,实现了不同地点、不同实验室、不同检测机构采集的食品安全检验检测数据的高效分析及价值挖掘,是一种基于云计算下海量数据信息为中心的风险交流方法。随着数据量的增多,数据挖掘结果的准确性会更加有所提高,分析结果也会更加有价值,该风险分析算法的优势会愈发明显。

参考文献

- [1] ALBERT R, JEONG H. Internet: Diameter of the World - Wide Web[J]. Nature, 1999, 401(6):130-131.
- [2] MELL P M, GRANCE T. The NIST Definition of Cloud Computing: SP 800-145[M]. Washington, DC: National Institute of Standards & Technology, 2011.
- [3] 谢世清. 论云计算及其在金融领域中的应用[J]. 金融与经济, 2010(11):9-11.
- [4] 杨大泉. 云计算下的金融信息集成与服务:兼论上海证劵报战略转型规划[D]. 上海:上海交通大学, 2013.
- [5] 肖四如,肖可砥. 云计算时代的金融创新[J]. 银行家, 2013(3):116-119.
- [6] 曹子健,林今,宋永华. 主动配电网中云计算资源的优化配置模型[J]. 中国电机工程学报, 2014, 34(19):3043-3049.