

doi:10.11751/ISSN.1002-1280.2017.10.09

兽药大数据平台的应用架构研究

高录军,刘玲,张积慧,郝毫刚,李晓平,唐军,刘雨,赵丽丹*

(中国兽医药品监察所,北京 100081)

[收稿日期] 2017-07-26 [文献标识码] A [文章编号] 1002-1280 (2017) 10-0062-06 [中图分类号] S851.66

[摘要] 兽药大数据对兽药监管、兽药行业健康发展以及疫病防控具有重要意义。针对现有的兽药海量数据进行大数据应用研究,提出了兽药大数据平台的应用方向和应用架构,以为兽药大数据平台的应用发展提供参考。

[关键词] 兽药;大数据;应用;模式

Research on Application Framework of Big Data Platform for Veterinary Drugs

GAO Lu-jun, LIU Ling, ZHANG Ji-hui, HAO Hao-gang, LI Xiao-ping, TANG Jun, LIU Yu, ZHAO Li-dan*

(China Institute of Veterinary Drug Control, Beijing 100081, China)

Corresponding author: ZHAO Li-dan, E-mail: zhaolidan@ivdc.org.cn

Abstract: The veterinary drug big data is in great significance towards the veterinary drug supervision, healthy development of the industry, and disease prevention and control. In this paper, the application direction and structure of the big data platform for veterinary drugs are put forward. This can provide a reference for the application and development of the big data platform for veterinary drugs.

Key words: veterinary drug; big data; application; mode

基金项目: 农业部重大信息平台构建及运维专项(2130104)

作者简介: 高录军, 硕士, 工程师, 从事兽药信息化建设及研究。

通讯作者: 赵丽丹。E-mail: zhaolidan@ivdc.org.cn

[17] 吴文博, 董占军. 中药制剂中细胞破壁技术探讨[J]. 中国药房, 2011, 22(3):285-287.

Wu W B, Dong Z J. Exploration of cell wall broken technology of traditional Chinese medicine preparation [J]. China pharmacy, 2011, 22(3):285-287.

[18] Wang D, Fan Z M, Deng L R, et al. Review on classification and application prospect of ultrafine powder material [C]. Materials Science Forum. 2014:272-277.

[19] Eitzler F M, Uddin M N. Powder technology and pharmaceutical

development: particle size and particle adhesion [J]. Kona, 2013, 30(30):125-143.

[20] 蔡璐, 梁少瑜, 戴开金, 等. 人参超微粉与细粉的体外溶出度比较[J]. 南方医科大学学报, 2013, (10):1547-1550.

Cai L, Liang S Y, Dai K J, et al. Comparison of dissolution properties between Ginseng micropowder and common powder [J]. J South Med Univ, 2013, (10):1547-1550

(编辑:陈希)

随着互联网、物联网和云计算技术的迅猛发展,形成的海量数据对人类社会产生日益重大的影响。与此同时,数据也成为一种新的自然资源^[1],亟待人们对其加以合理、高效、充分的利用,使之能够给人们的生活工作带来更大的效益和价值^[2]。为了将数据转化为知识,可以利用数据仓库、数据挖掘等技术对数据进行处理与分析,从而辅助相关机构做出决策参考,提高决策能力、决策效率和决策准确性^[3]。但是,收集到的数据越来越多、数据结构越来越复杂,一般的数据挖掘技术已经不能满足需要,这就使得“大数据”应运而生。相对数据挖掘而言,大数据的应用不仅仅是数据量级的差别,也是对数据应用层次的提升,不仅仅对传统的规则数据进行处理,同时,也对杂乱的非结构化数据进行分析,从而对数据内在的知识进行获取,从而支撑相关领域的预测、规划和决策。

随着兽药信息化的不断发展,国家兽药产品追溯系统^[4]、国家兽药基础信息查询系统、各省级兽药监管平台、兽药实验室管理系统、兽药企业信息系统等兽药信息系统的建设,为兽药行业提供了海量的数据和资源,为兽药大数据的应用奠定了基础。因此,开展兽药大数据平台的应用架构研究能够充分利用大量的兽药信息化数据,为兽药信息化的深入发展和兽药行业的健康发展提供有力支撑。

1 兽药大数据平台的功能

1.1 功能概述 兽药大数据平台建设以兽药行业基本发展情况分析、兽药行业发展趋势分析、兽药企业发展分析、兽药行业预测分析等为主要功能。兽药行业基本发展情况分析从兽药生产、经营、使用三个维度分析兽药的生产规模、流通水平及使用情况。行业发展趋势分析从兽药行业产品结构入手,分析现有兽药及新兽药研发的发展趋势,同时根据兽药结构,分析兽药生产产能及产能分布情况,对产能进行同比环比的分析,探明兽药生产增长趋势,及时根据产能数据变化,规避行业发展的风险。兽药企业发展分析通过大数据,为企业提供分析报告,以帮助企业从各方面提升产品质量、提高市场竞争力。兽药行业预测分析通过大数据分

析,对兽药行业监管、用药安全及疾病预防方面提供预测分析。

1.2 基本功能

1.2.1 兽药行业基本情况分析 兽药行业基本情况分析包括兽药生产企业分析、兽药经营企业分析、兽药使用企业分析以及兽药产品分析等。兽药生产企业分析包括兽药生产企业的数量及变化情况、区域企业数据及变化情况、企业规模分析(包括产值规模、资产规模、人员规模等)、经济效益分析。兽药经营企业分析包括兽药经营企业整体数据及变化情况、区域企业数量及变化情况。兽药使用企业分析包括兽药使用企业整体数量及变化情况、区域企业数量及变化情况。另外,同时三个维度进行兽药产品的深层次分析,包括兽药进出口情况、使用量变化、使用种类变化、区域使用情况等。

1.2.2 兽药行业发展趋势分析 兽药行业发展趋势分析主要包括兽药产品结构分析、兽药产能数据分析及兽药流通数据分析。兽药产品结构分析主要对不同种类和不同分类的兽药进行结构分析,比如中药、化药、生物制品等不同药品的适应症进行分析,同时,对新兽药研发情况进行分析。兽药产能数据分析主要根据兽药分类信息,对各类药品的产能分布及增长趋势进行分析,按照时间、区域等情况进行分析,为行业投资提供依据。兽药流通数据分析主要从市场容量、流通效率及流通覆盖情况,分析行业整体兽药流通情况,发现兽药流通规律、有效规避兽药流通风险等。

1.2.3 兽药企业发展分析 兽药企业发展分析主要是通过兽药大数据,为企业提供宏观及微观分析报告,以帮助企业从各方面提升产品质量、提高市场竞争力。包括兽药产品结构、市场覆盖范围、市场占有率、产品流通效率、销售渠道、产品库存、产品流向、企业预警及其它预测分析。

1.2.4 兽药行业预测分析 通过兽药大数据分析,对兽药行业监管、药品用药安全及疾病预防方面提供预测分析,为兽药监管和动物疾病防控提供决策参考。兽药行业预测分析主要包括兽药流通异常预警分析、过期兽药预警分析和疾病预测分析等。

兽药流通异常预警分析主要根据大量兽药流通数据,通过同比、环比的方式,及时发现兽药流通的异常状况,及早发现兽药流通的异常状况。过期兽药预警分析主要根据兽药的生产日期及兽药的流通数据,及时对市场上的过期兽药进行事先预警,借助国家兽药产品追溯码信息,对过期兽药进行及时

预警。疾病预测分析根据兽药的生产数据、流通数据及兽药使用数据,对一定时间段之内的疾病情况进行分析;根据历史同期用药数据,预测可能发生的疫情情况,并为相关单位提供预警报告。

1.3 功能结构 根据兽药大数据平台相关功能,提出平台功能结构(图 1)。

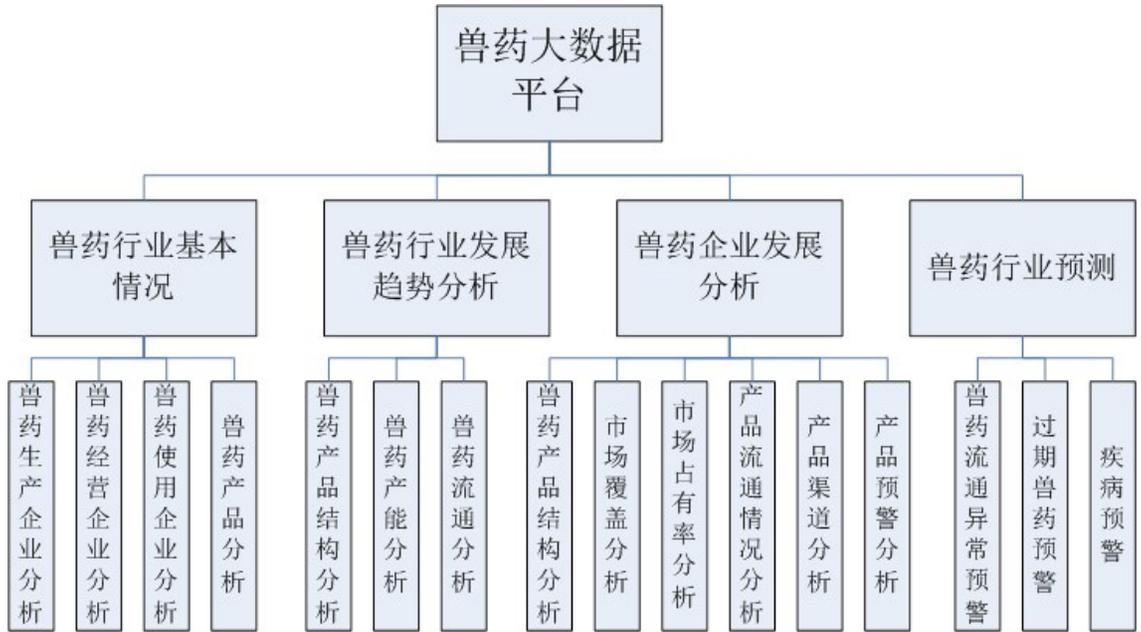


图 1 兽药大数据平台功能图

Fig 1 Function of the veterinary drug big data

2 兽药大数据处理

2.1 数据源 兽药大数据主要来源于兽药信息相关的数据,包括结构化数据、半结构化数据和非结构化数据。

结构化数据也称作行数据,通常为关系型数据库中存储的数据,比如 ORACLE、SqlServer 等,是由二维表结构来逻辑表达和实现的数据,严格地遵循数据格式与长度规范,主要通过关系型数据库进行存储和管理。结构化数据主要来源于国家兽药产品追溯系统、国家兽药基础数据查询系统、省级兽药管理平台等多个兽药信息化系统及其他相关来源。其中,国家兽药产品追溯系统兽药追溯数据约 500 亿条,国家兽药基础数据查询系统数据约 30 万条,省级兽药管理平台数据约 10 亿条,以及其他兽药信息系统的大量数据。

半结构化数据是相对于数据库中严格的二维表结构而言的,是一种标记信息,比如 XML 等。半结构化数据主要来源于国家兽药产品追溯系统数据采集过程中所产生的 XML 文本信息等。

非结构化数据是数据结构不规则或不完整,没有预定义的数据模型,不方使用数据库二维逻辑表来表现的数据。包括所有格式的办公文档、文本、图片、HTML、视频信息等等。非结构化数据主要来源于所有兽药相关的文档、中国兽药信息网等网页信息。

2.2 数据处理流程 按照大数据处理的一般流程^[5],兽药大数据平台可以分为数据采集、数据处理、数据分析和数据解释四个阶段。

数据采集主要是通过传感器、条码识别、搜索引擎以及数据检索工具等对数据源进行获取,从而

获取兽药大数据平台的基础数据。数据处理主要是对采集的数据进行关联与聚合,对数据进行清洗,从而产生可以信赖的数据。数据分析主要是通过数据挖掘、机器学习和统计分析的工具与方法对数据进行深层次的分析,从而形成一定的结论。数

据解释主要是对数据分析后的结果通过可视化展现,能够让用户快速理解结论。

按照大数据处理一般流程,兽药大数据平台的数据处理流程见图 2。

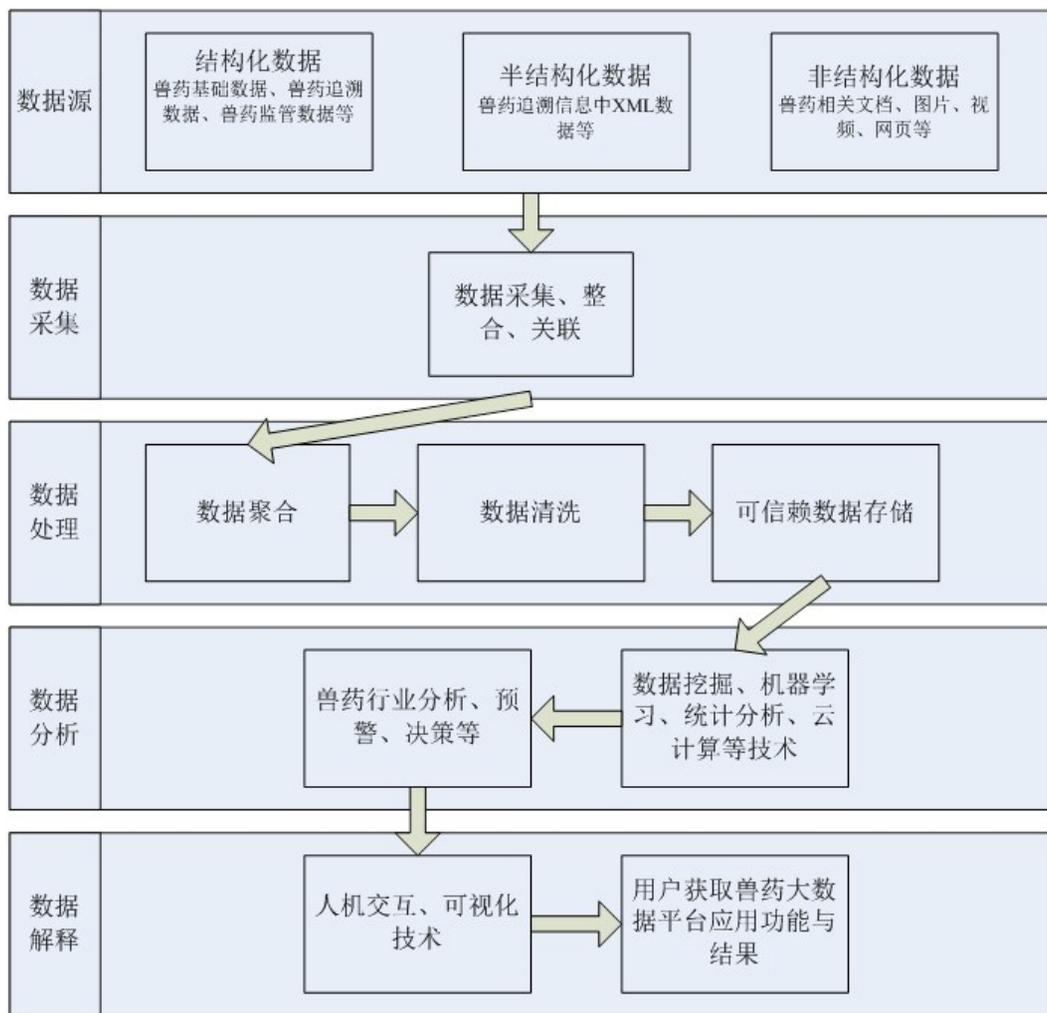


图 2 数据处理流程图

Fig 2 Data processing

3 兽药大数据平台架构

兽药大数据平台根据数据处理流程,其架构共分为五个层次,自底向上分别为基础层、数据源层、数据处理层、大数据层、应用层。

基础层为兽药信息化及兽药大数据平台的物理设备支持,包括系统运转的服务器、数据库、操作系统、网络、中间件以及数据采集设备等。

数据源层主要为兽药大数据平台的基础数据来源,包括结构化数据(如国家兽药产品追溯系统数据库、国家兽药基础数据查询数据库等)、半结构化数据(如国家兽药产品追溯系统采集数据等)、非结构化数据(如兽药信息相关的文档、图片、视频及网页等)。

数据处理层主要是根据数据的结构特征等按

照大数据平台设计开发适配器,对数据进行清洗、转换、处理并最终存入大数据平台的数据库。

大数据层是兽药大数据平台的核心,可以采用 MapReduce 分布式计算技术作为编程模型和计算框架,为系统提供高性能的分布式计算环境。大数据存储可采用基于列存储的非关系型数据库 Hbase, Hbase 是一种建立在 HDFS 之上、实时读写、

高性能的非结构化的分布式数据库系统。

应用层分为数据分析层和可视化层。数据分析层主要是通过数据挖掘、机器学习和统计分析等工具和方法对相关数据进行深入分析,形成分析结果。可视化层是将系统分析结果以可视化的形式予以展示,提供给用户,为科学决策提供参考。

兽药大数据平台的架构如图 3 所示:



图 3 兽药大数据平台架构图

Fig 3 Architecture of veterinary drug big data

4 结 语

兽药大数据平台建设是兽药管理和兽药信息化发展的必然趋势,兽药大数据平台的应用架构研究能够充分利用海量兽药数据,将兽药数据转化为知识,能够促进兽药监管的便捷化和科学化,能够促进兽药企业科学生产,提高企业生产效率,为兽药行业和兽药企业的健康发展提供有力手段和有效支撑。同时,能够帮助相关部门有效预判、预测动物疫病并提前做好充分准备,为保障动物养殖安全和食品质量安全起到积极的作用。

参考文献:

- [1] 于艳华,宋美娜. 大数据[J]. 中兴通讯技术, 2013(1): 57-60.
Yu Y H, SONG M N. Big data[J]. Zte Technology Journal, 2013(1):57-60.
- [2] 刘智慧,张泉灵. 大数据技术研究综述[J]. 浙江大学学报, 2014, 48(6): 957-972.

Liu Z H, ZHANG L Q. Research overview of big data technology [J]. Journal of Zhejiang University, 2014, 48(6): 957-972.

- [3] 熊忠阳. 面向商业智能的并行数据挖掘技术及应用研究[D]. 重庆:重庆大学,2004.
Xiong Z Y. Research on palallel data mining and application for business intelligence [D]. Chongqing: Chongqing University, 2004.
- [4] 高录军,刘业兵,李晓平,等. 国家兽药追溯系统信息采集与处理技术的研究[J]. 中国兽药杂志, 2014, 48(9): 53-56.
Gao L J, Liu Y B, Li X P, *et al.* Research on information acquisition and processing of national veterinary drug product traceability information system [J]. Chinese Journal of Veterinary Drug, 2014, 48(9): 53-56.
- [5] 孟小峰,慈祥. 大数据管理:概念、技术与挑战[J]. 计算机研究与发展, 2013,50(1):146-169.
Meng X F, Ci X. Big data management: concepts, techniques and challenges[J]. Journal of Computer Research and Development, 2013,50(1): 146-169.

(编辑:李文平)