食品大数据应用综述

陶 启[1,2]，李 伟[3,4]，丁红卫[1,2]，崔晓晖[*1,2]

（1. 武汉大学 空天信息安全与可信计算教育部重点实验室，湖北 武汉 430072；2. 武汉大学 国家网络安全学院，湖北 武汉 430072；3. 江南大学 人工智能与计算机学院，江苏 无锡 214122；4. 江南大学 江苏省媒体设计与软件技术重点实验室，江苏 无锡 214122）

摘要：大数据技术是指从各类数据中快速获得有价值信息的能力。大数据技术方兴未艾，为食品科学提供了一种全新的思维方式和分析方法，可以提升食品科学在食品追踪溯源、食物配对、食谱推荐、食品安全、食品生产管控、用户食品偏好、未来食品分析上的预测性以及准确性，然而在大数据应用系统中存在数据缺失、难以辨真伪等挑战。本文中阐述了食品大数据的应用现状，分析了其面临的挑战并讨论可行性解决方案，展望了食品大数据未来发展趋势。本文作者的目的是希望食品大数据领域研究者在本文研究基础上，继续从大数据角度为食品科学添砖加瓦，并根据未来食品发展趋势，实现食品科学提升。

关键词：食品大数据；食品追踪溯源；食物配对；食品安全；食品生产管控

中图分类号：TP 399　**文章编号：**1673-1689(2020)12-0001-05　**DOI：**10.3969/j.issn. 1673-1689.2020.12.001

A Review of Big Data Application in Food Industry

TAO Qi[1,2]，*LI Wei*[3,4]，*DING Hongwei*[1,2]，*CUI Xiaohui*[*1,2]

(1. Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, Wuhan University, Wuhan 430072, China; 2. School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China; 3. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; 4. Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China)

Abstract: Big data technology refers to the ability to quickly obtain valuable information from various types of data. Big data technology is developing to make food science serve people better, which provides a new success insight and a scientific analysis strategy for food science, which can improve the performance of food science in food tracing, food matching, recipe recommendation, food safety, food producing management, food preference, and predictability and accuracy of food analysis in future. However, there are some challenges in big data application, such as data loss and difficulty to distinguish authenticity. This review introduced the application of big data in food industry, analyzed the challenges and the feasible solutions, and prospect tendency of food big data. Food scientists in the research area of food big data are encouraged to do further study based on our survey, contributing to food science from the perspective of big data and realizing the improvement of food science according to the future development trend of food.

Keywords: food big data, food tracing, food matching, food safety, food producing management

收稿日期：2020-06-01

基金项目：国家“十三五”重点研发计划项目(2018YFC1604000)；中央高校基本科研项目(2042017gf0035，2042019kf0268)。

作者简介：陶　启(1992—)，男，博士，主要从事区块链、大数据与食品安全方面研究。E-mail：qtao17@whu.edu.cn

*** 通信作者：**崔晓晖(1971—)，男，博士，教授，博士研究生导师，主要从事大数据、区块链与食品安全研究。E-mail：xcui@whu.edu.cn

随着计算机的发展,数据采集比以往任何时候更加便利快捷,由此带来大量的、各种类型的数据。随着大数据技术的兴起,从数据中通过分析获取多智能的、深入的、有价值的信息正深刻影响着现实生活中的各行各业,尤其是对于食品行业。当前,我们国家对于食品要求已经从"吃的饱"转变为"吃的健康",而传统的食品科学技术已不能满足人们日益增长的食品要求。因此,将大数据技术应用到食品科学领域,为食品科学提供了一种全新的思维方式和科学分析方法,可以提升食品科学在食品追踪溯源、食物配对、食谱推荐、食品安全,食品生产管控、用户食品偏好、未来食品分析上的预测性以及准确性,使得食品科学更好地为人们服务。

作者首先阐述食品大数据的应用现状,然后分析目前食品大数据面临的挑战,最后总结并展望食品大数据的未来发展趋势。

1 食品大数据应用

食品行业是当今最具活力的行业之一,产品必须随着消费者的需求而不断发展。食品大数据已在食品行业获得广泛应用,对消费者饮食健康、食品安全、行业拓展等领域影响较大,为行业创造了巨大经济价值与社会效益。下面分别从食物配对与风味设计、食谱推荐、食品安全管控、食品市场开拓、食品外卖应用和打造企业品牌 6 方面对食品大数据的行业应用进行概述。

1.1 食物配对与风味设计

1992 年,厨师 Heston Blumenthal 和 Francois Benzi 认为具有相似风味的原料烹调效果会很好,于是有厨师就实现了白巧克力和鱼子酱的风味组合,但由于白巧克力和鱼子酱的化学性质差异大,导致组合后的食物味道非常特殊。随着食品科学的发展,我们现在掌握了大量的食物化学物质信息[1-4],这些物质使得不同的食物具有不同的气味和味道。有了这些信息,Ahn 和 Ahnert 等学者[5-6]开发了一个特色风味网络(flavor network),一个由共享的风味化合物连接的原料网络。由于 Ahn 等学者收集的原材料有限,因此组合后的食材风味也有限。鉴于此,Garg 等学者开发了 FlavorDB 数据库[7],FlavorDB 数据库比 Ahn 等学者收集的食材更为丰富,但是其菜品组合网络仍然是基于 flavor network。学者 Simas 等提升了 flavor network,并将其设计的网络称之为 Food-bridging[8]。以上方法都是基于类似化学成分组合食物,很多著名的食物搭配(比如红酒和牛肉)并没有共同的化学成分或者风味化合物,但是搭配一起仍然广受人们喜爱。因此需要在更大范围内寻找食物配对,而不是仅仅基于风味化合物或者化学成分。

在未来食品风味设计方面,香料公司 McCormick 在 2019 年与 IBM 合作,利用人工智能与大数据预测新的风味组合。通过分析来自数百万个数据源的数据以改进该公司的"风味预测"系统。该系统自动生成新的风味特征并吸引了消费者的关注。

1.2 食谱推荐

食谱推荐(recipe recommendation)也是食品大数据分析领域中常见的应用场景。Teng 等学者提出利用食谱推荐算法来确定食物成分是否在菜谱中为必需品[9]。Grace 和 Maher 等学者结合基于案例(case-based)的推理和深度学习算法来生成新的菜谱[10-11]。但是,由于深度学习的不确定性,后者可能会生成"黑暗料理"。另外,在食谱推荐领域,有学者专门聚焦于"食疗"。Freyne 等学者针对肥胖症来设计食谱推荐[12]。他们根据医学专业人士的建议和对肥胖人士的调研,设计旨在给肥胖症人群个性化推荐食物。Yoshida 等学者则聚焦于用户食物偏好上,这是因为上述设计的食物推荐系统训练数据来源于开放的食谱网站[13]。这些食谱网站的特点是访问频繁,网站储存的大部分食谱非常大众化,部分人群则有自己的食物偏好。他们利用人们的食谱浏览和烹饪历史数据,提出了一种基于用户食物偏好的个性化食谱推荐系统,用以满足人们的食物偏好。

1.3 食品安全管控

1976 年联合国卫生组织(WHO)构建的全球环境监测系统(GEMS/Food),由各参与机构提交食品污染物浓度数据组建数据中心,结合数据处理与分析技术,帮助各国政府、CAC 等机构评估食品污染物污染等级与发展趋势[14],是各国用于评估与管理食源性风险的重要工具。2015 年世界卫生组织(WHO)整合农业、食品、公共卫生和经济指数等领域多元化数据构建食品安全大数据服务平台 FOSCOLLAB,通过增强采集数据源维度提升对食品安全风险监控效果[15]。2014 年贵州科学院整合省内各级政府、检测机构及入驻企业的产品数据,建立

食品安全与营养大数据云平台，实现食品防伪验证、安全风险监测与预警[16]。

食品追溯系统中存在的数据关联性弱、数据记录不准确等问题，为食品召回与追溯问题源头带来挑战。Badia-Melis 等提出食品追溯本体应用(FTTO)模型，基于语意分析对异构数据库数据进行整合建立归一化的知识体系，基于唯一标识原则对整合后数据进行统一编码处理，为溯源数据的电子交换提供一个国际、非专属性统一标准[17]。

1.4　食品市场开拓

在食品市场中，产业链内外都蕴藏着庞大的食品相关数据，通过收集与分析能够促进企业开拓市场[18]：1)通过收集商品及其零售信息进行分析能够了解市场近况，把握竞争者的商业动态，明确产品的市场定位，从而掌握市场先机；2)通过收集消费者消费信息建立客户大数据库，通过对客户进行用户画像获取消费者的喜好、消费倾向、价值倾向及当前商品的市场口碑状况，为企业制定高效的营销策略提供科学指导建议，为消费者提供优质服务赢得消费者信赖，让企业在市场竞争环境下能够不断壮大；3)通过收集社交平台中各种食品的各类点评信息组建产品需求大数据，通过聚类、统计分析，发掘新的消费需求与产品的潜在质量问题，从而改进或研发新产品、量化价值并制订合理产品价格提升服务质量。

1.5　食品外卖应用

我国线上外卖用户量占网民总量 44%以上，规模已超过 3.98 亿，且将继续保持增长趋势，外卖未来具有广阔市场前景[19]。用户体量大、高速增长的外卖市场产生了庞大的外卖数据，外卖大数据服务平台的建设与应用将是未来发展趋势，不仅有助于帮助政府对外卖餐饮业进行监管，而且能够为外卖餐饮业创造巨大经济价值与社会价值，包括：1) 通过准确预测并告知客户外卖送达时间，能够避免对消费者日常计划的影响，帮助企业树立良好品牌形象；2)帮助食品及其相关行业了解项目或商品的流行趋势；3) 外卖大数据应用平台促使外卖供应链的透明化，供应链的透明性有利于建立或改善客户与外卖企业间的信任关系，同时有助于督促企业为客户提供更高品质的产品及更优质的餐饮服务；4) 通过分析外卖数据集能够清晰了解城市整体运行状态[20]。

由于外卖数据直接或间接涉及客户地理位置、偏好、银行、身份、通信等敏感隐私信息，一旦泄露将对客户财产及生命健康带来安全隐患，因此外卖大数据平台构建将面临更为严格的大数据安全、效率等方面挑战。

1.6　打造企业品牌

大数据技术在食品行业中的推广普及，让食品产业链中企业能够更好地了解上下游客户需求，预测食品行业发展趋势，营造企业品牌形象。可采取措施：1)基于市场分析结合客户的购买历史预测其下一步可能购买的产品，利用这些信息，食品企业可以制定优惠价格及商品组合，以提高客户满意度并确保重复业务；2)通过收集社交媒体中用户对于食品的有关评价信息进行分词、聚类及情感分析，建立积极、消极或中立的态度，通过跟踪这些不断变化的行为和喜好，可以让食品企业在发布负面新闻或不良反馈时采取行动；3)结合人工智能，大数据为企业提供了巨大的增长机会。例如，以多传感器信息融合为基础，将机器视觉应用于保健酒的缺陷检测中，可以实现外观缺陷和可见异物的不合格产品检测[21]；4)大数据让食品产业链更加透明，不仅能够实现产品的全程监管，而且能够以消费者众包模式来监督食品品质、服务质量及食源性疾病或食品事件的爆发趋势。

2　食品大数据挑战

目前食品大数据应用系统中存在数据缺失、数据量不足、难以辨真伪等挑战，研究食品大数据高效采集、可信校验与跨部门共享具有重要意义。

大数据发展的前提条件是要有丰富的数据源，即使对于数据化程度比较高的 IT 行业，依然缺少资源共享和信息交换机制，只能在企业内部探索和尝试。政府部门缺乏数据开放的动力，由于其掌握的数据有一定的敏感性，因此趋于保守态度。各大企业不会随便开放自身有价值的数据，因为它有巨大的商业价值，也关系到企业的生死存亡。长此以往，不同的部门数据源独立存在，不能够互相共享，形成了一个个数据孤岛，无法实现行业跨部门全链条大数据体系。大数据困境从早期的难以处理大数据到今天变为难以得到全链条大数据。为解决该问题，政府期望采取强制手段，建立数据中心，整合管辖范围的数据，但数据隐私保护难以解决。阿里、腾

讯等互联网巨头期望通过并购来构筑庞大的行业帝国实现行业全链条数据企业内部化。但行业数据垄断的大企业会利用自身垄断地位阻碍创新,使垄断地位更加坚固。

食品大数据覆盖从“农田到餐桌”全过程,涉及到的部门众多。这种由于数据保密和隐私保护等造成的各数据拥有方不愿共享数据依然是食品大数据有效运用的主要障碍。需要提出一种新型的去中心化数据采集、存储、共享及智能分析机制来解决目前面临的挑战。从目前来看,“区块链+大数据”具备很大的潜力。区块链是一个公共账本、一种集体维护的网上数据库, 与传统数据库的主要区别在于:匿名性、去中心化、不可篡改、分布式存储、多备份、隐私保护等。从技术的角度来看,我们经常说“区块链技术”,其实并不是一种单一的技术,而是多种技术整合的结果。这些技术以新的结构组合在一起,形成了一种新的数据记录、存储和表达方式。但区块链也存在新型基于区块链–云计算的去中心化数据采集方法与存储机制,允许数据拥有方将原始数据保存在私有云服务器端, 实现链下存储,同时抽提少量数字水印和属性数据,通过区块链进行链上存储。通过在区块链中设置校验机制,保证链下存储信息的真实性及防篡改;通过基于区块链技术的多方计算架构,其他用户可以在不直接访问数据拥有方原始数据情况下运算数据,从而实现对数据的私密性进行保护,杜绝数据共享中的信息安全问题。解决全链条跨部门数据采集融合困难,数据孤岛、完整性与可信度差的难题。

3 展 望

随着科技的发展,食品行业累积了大量、来源广泛、增长速度快、价值密度低、应用价值大的数据。如何使用大数据技术挖掘出食品数据中的潜在应用价值和促进食品行业的可持续发展,已经成为食品领域重点研究的问题。食品大数据已经在食物配对与风味设计、食谱推荐、安全管控、市场开拓、外卖应用和打造企业品牌等领域获得良好应用,促进了行业经济发展。针对目前食品大数据应用中存在的数据孤岛、难辨真伪等问题,“区块链+大数据”是解决数据拥有者间的数据共享、促进食品大数据智能分析应用的可行方案。面对日益剧增的粮食需求与有限资源之间矛盾凸显,进一步利用大数据人工智能等信息技术降低资源消耗并研发新一代食品生产技术将是未来食品科研工作者的重点方向之一。

资源浪费及食源性疾病暴发是制约食品产业健康可持续发展的关键因素,具体包括:1)食品种养植过程滥用化肥、农药等,致使农产品中农残超标、水土污染严重等问题[98];2)发展中国家易腐食品损失严重,发达国家消费者浪费的食物总量巨大[22];3)食品加工及运输过程中能耗高、环境污染大;4)由于无法快速有效溯源食品污染源头,通常将受影响食品作为病原体的潜在来源而全部丢弃。物联网、大数据、区块链等信息技术将是促进食品产业可持续发展的核心要素。物联网能够部署贯穿食品产业链的各个环节并采集数据。通过对物联网采集到的种植大数据与农田地理大数据进行综合分析实现精准农业,构建优质、高产、高效的农业生产模式,提高种植户经济效益、降低资源浪费与水土污染;提高食品产业链透明度能够减少食品产业链加工过程浪费、减少消费者的食品浪费数量。区块链与大数据结合可以使得食品供应链完全透明,实现:1)优化产业链结构减少加工浪费,优化产品调度策略降低运输能耗,减少尾气排放;2)食源性疾病暴发时,实现污染源精准溯源及污染品的快速找回,保障食品供应链上下游企业效益与品牌形象[23]。基于大数据画像技术对消费者的消费、饮食等数据进行分析,为消费者提供健康膳食建议,提升消费者生活品质,降低营养过剩食品损耗。

至 2050 年,全球人口数量将超 90 亿,面对有限的资源,基于大数据、人工智能的未来食品开发也将提速增效。未来食品的研发需要融合食品生物、食品大数据、组学大数据、深度学习医疗健康等学科知识,开发出健康美味、风味独特、营养均衡的新一代食品,满足日益增长的人口需求。我国要把握未来食品研究进入生物组学大数据时代的重要战略机遇,尽早布局,充分融合食品分子分析、电子信息、数据建模和分析等技术,根据食品营养组分和芳香分子特性, 通过食物配对和食品风味搭配,设计开发新一代食品。

参考文献:

[1] TAJKARIMI M M,LBRAHIM S A,CLIVER D O. Antimicrobial herb and spice compounds in food[J]. **Food Control**,2010,21(9):1199-1218.
[2] GUICHARD E. Interactions between flavor compounds and food ingredients and their influence on flavor perception[J]. **Food Reviews International**,2002,18(1):49-70.
[3] MAGA J A,KATZ I. Simple phenol and phenolic compounds in food flavor[J]. **Critical Reviews in Food Science & Nutrition**,1978,10(4):323-372.
[4] VARA-UBOL S,CHAMBERS E,CHAMBERS D H. Sensory characteristics of chemical compounds potentially associated with beany aroma in foods[J]. **Journal of Sensory Studies**,2004,19(1):15-26.
[5] AHN Y Y,AHNERT S,BAGROW J,et al. Flavor network and the principles of food pairing[J]. **Scientific Reports**,2011,1:196.
[6] AHN Y Y,AHNERT S. The flavor network[J]. **Leonardo**,2013,46(3):272-273.
[7] NEELANSH G,APUROOP S,RUDRAKSH T,et al. FlavorDB:a database of flavor molecules[J]. **Nucleic Acids Research**,2017,46(1):1210-1216.
[8] SIMAS T,FICEK M,DIAZ-GUILERA A,et al. Food-bridging:a new network construction to unveil the principles of cooking[J]. **Frontiers in ICT**,2017(4):1-18.
[9] TENG C Y,LIN Y R,Lada A Adamic. Recipe recommendation using ingredient networks [C]//In Proceedings of the 4th Annual ACM Web Science Conference,[S.l.]:ACM,2012:298-307.
[10] GRACE K,MAHER M L,WILSON D C,et al. Combining CBR and deep learning to generate surprising recipe designs[C]//In International Conference on Case-based Reasoning. Cham:Springer,2016:154-169.
[11] GRACE K,MAHER M L. Surprise-triggered reformulation of design goals [C]//In 13th AAAI Conference on Artificial Intelligence. Phoenix:AAAI,2016:3726-3732.
[12] FREYNE J,BERKOVSKY S. Intelligent food planning:personalized recipe recommendation [C]//Proceeding of the 15th international conference on Intelligent user interfaces. Hong Kong:ACM,2010:321-324.
[13] UEDA M,TAKAHATA M,NAKAJIMA S. User's food preference extraction for personalized cooking recipe recommendation [C]//International Semantic Web Conference. Boston:IEEE,2011:98-105.
[14] Moy G G . GEMS/food and total diet studies[M]. New York:Springer,2013.
[15] WORLD HEALTH ORGANIZATION. FOSCOLLAB:Global platform for food safety data and information [EB/OL]. (2015-01-12)[2020-04-20]. https://www.who.int/foodsafety/foscollab/en/
[16] 刘彤,谭红,张经华. 基于大数据的食品安全与营养云平台服务模式研究[J]. 食品安全质量检测学报,2015,6(1):366-371.
[17] BADIA-MELIS R,MISHRA P,RUIZ-GARCÍA L. Food traceability:new trends and recent advances. a review [J]. **Food Control**,2015(57):393-401.
[18] 李静. 基于大数据精准营销的网络营销策略研究[J]. 商业经济研究,2017(11):46-47.
[19] 前瞻产业研究院. 2020-2025 年中国在线外卖商业模式与投资战略规划分析报告[EB/OL]. (2020-05-13) [2020-05-20]. https://www.qianzhan.com/analyst/detail/220/200512-65621d53.html
[20] 章胜. 396 万外卖大数据见证“武汉醒来”[N]. 长江日报,2020-04-10(015).
[21] 孟庆宇,蒋甜燕,孔伟,等. 浅谈机器视觉检测在预包装食品加工过程中的应用[J]. 化工管理,2019(32):50-52.
[22] PARFITT J,BARTHEL M,MACNAUGHTON Food waste within food supply chains:quantification and potential for change to 2050[J]. **Philosophical Transactions of the Royal Society B–Biological Sciences**,2010,365(1554):3065-3081.
[23] KAMATH R. Food traceability on blockchain:Walmart's pork and mango pilots with IBM[J]. **The Journal of the British Blockchain Association**,2018,1(1):3712.