

基于 LDA 模型的大兴区都市农业发展方向研究

皮婧文^{1,2}, 王海燕^{1,2}, 吴保国^{1,2}

(1. 北京林业大学信息学院, 北京 100083;

2. 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083)

摘要 在全面实现农业现代化的冲刺阶段, 大兴区都市农业的发展对于区域实现农业现代化具有重要影响。文章首先从大兴区都市农业发展的现状、遇到的问题、面临的机遇展开分析; 然后运用 LDA 主题模型对北京大兴野生动物园的游客评论进行主题分类, 并根据实验结果探索大兴区都市农业三产融合的发展方向; 最后从优化大兴区传统农业结构、深度开发农耕文化和完善“大兴旅游”APP 三个角度提出发展对策。

关键词 都市农业 发展方向 LDA 主题模型

“十三五”时期是我国全面建成小康社会的决胜阶段, 也是全面实现农业现代化的冲刺阶段。农村小康是全面建成小康社会中最艰巨最繁重的任务, 为使发展成果更多更公平地惠及全体人民, 国家强农惠农的政策力度势必不断加大; 同时, 随着市场配置资源力度进一步强化, 社会资本不断向农业领域渗透, 农业的发展潜力将得到进一步释放^[1]。在此之前, 很多学者已经对北京市都市农业发展模式和发展对策展开了研究^[2-3]。而在“十三五”期间, 大兴区都市农业的发展将会面临更多机遇和挑战, 文章将通过分析大兴区都市农业发展的现状、遇到的问题、面临的机遇, 迎接挑战, 探索适合大兴区的都市农业发展方向。

1 都市农业发展现状与发展问题

早在 20 世纪末, 北京市委、市政府就提出了以“精品农业”、“设施农业”、“观光农业”、“创汇农业”、“籽种农业”、“加工农业”等 6 种农业为切入点的都市农业发展思路^[4]。大兴区都市农业在发展之初, 形成了以设施农业、畜牧种业和休闲农业为代表的发展模式。近几年受平原造林工程、城镇化建设、非首都功能疏解以及农业绿色发展的影响, 大兴区都市农业的发展面临不少挑战。

1.1 设施农业方面

(1) 设施农业的发展现状: 大兴区作为北京市的“南菜园”, 设施农业生产一直都是农民增收的主要途径之一。由图 1 可知, 近年来, 大兴区设

施农业收入占北京市设施农业收入的比例一直保持在较高水平。此外还可以发现大兴区设施农业收入变化较大, 2008—2013 年大兴区设施农业收入整体呈稳定增加的趋势; 但是在 2014 年, 受拆迁、土地流转等因素影响, 播种面积连年下降, 收入也出现了小幅下降; 到了 2015—2018 年, 收入波动较小, 处于平稳期。

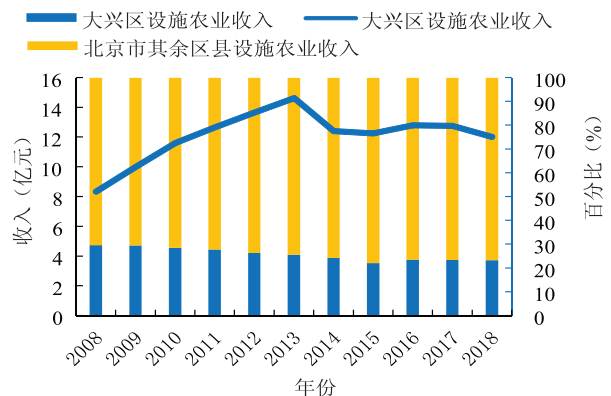


图 1 2008—2018 年大兴区设施农业收入占比北京市设施农业收入

施农业除了收入涨幅下降以外, 因农村青壮年劳动力资源不断向二、三产业转移, 且中小棚种植劳动强度大、效益低, 近年来中小棚种植面积正逐年减少。相对于中小棚种植面积的不断减少, 随着近年来社会资金不断注入农业生产经营, 各级政府和农户对农业投入的逐步加大, 大兴区设施农业标准不断提高, 温室和大棚面积占设施农业的比重

正逐年提升^[5]，由2013年不足七成的比重，增长至2018年超过九成的比重。

(2) 设施农业的发展问题：近几年，大兴区设施农业受拆迁、土地流转等因素影响，收入涨幅和播种面积连年下降。除此之外设施农业的发展还存在以下问题：①农业设施内出现种树行为。对于位于基本农田上的农业设施内种植林果业的行为要加快清理；位于一般农用地上的农业设施，要加快整治，严格按照设施农业政策恢复农业生产。②农业设施出现撂荒行为。对于撂荒的农业设施，要尽快组织进行生产；对于残破残缺日光温室、土棚温室、钢架大棚等已经不具备生产功能、存在安全隐患、影响农业生态环境的种养殖设施要全部清理。③农田和农业设施内仍存在住人现象，要尽快劝离仍居住在农田和农业设施内的人员。

1.2 种植业方面

(1) 种植业的发展现状：大兴区是京郊粮食主产区之一，受国家粮食直补政策影响，全区粮食生产面积曾一度达到2010年的1.43万 hm^2 。但近几年全区粮食生产面积逐年减少，到2018年，全区全年粮食收获面积和总产量同创历史新低，全年粮食实际播种面积为0.5万 hm^2 ，全年粮食总产量为3.1万t。全区从面积统计数据可知：2018年全区夏粮播种面积为0.11万 hm^2 ，秋粮播种面积为0.39万 hm^2 ，与2017年秋粮播种面积相比减少0.19万 hm^2 ，下降32.2%。从产量统计数据可知，2018年受夏、秋粮播种面积大幅下降影响，全区夏、秋粮总产分别创历史新低，夏粮实际总产量仅为0.56万t，比2017年减少0.34万t，下降37.8%；2018年秋粮实际总产为2.5万t，同比减少1.4万t，下降36.8%。

(2) 种植业的发展问题：在非首都功能疏解的大背景下，随着新农村建设和现代都市农业的发展，大兴区种植业的发展一方面因为生产规模缩减，导致所包含的粮食、蔬菜、花卉、瓜果等播种面积均呈现下降态势；另一方面种植业缺乏精品，种植技术需要加强，导致种植业生产发展整体呈下降态势^[6]，使得效益不断下降，出现持续萎缩，传统农业生产模式急需加快转型。

1.3 休闲农业方面

(1) 休闲农业的发展现状：随着农业结构的调整及农业高新技术的应用，大兴区各镇、村结合

自己的农业特点、自然资源，相继建成了一批高新农业观光园，为传统农业注入了一种全新的生产经营模式。大兴区休闲农业的发展由图2可知，2006—2013年其经营总收入总体呈增收趋势，但到了2014年经营总收入大幅回落降至1.32亿元，同比下降32.9%。2014年收入大幅回落一方面是受自然灾害影响，果品产量及采摘量骤减；另一方面是北京大兴国际机场建设方案落地，机场周边农业园将面临拆迁等问题，进而导致主要靠采摘收入支撑的农业观光园收入锐减。到了2015年经营收入有了小幅回升，但2016—2018年又呈现下降趋势。在收入下降期间，大兴区农业观光园总收入中门票收入、出售农产品收入和采摘收入占主导地位，3项收入共占观光园总收入比重超八成，多元化收入渠道有待进一步拓展。

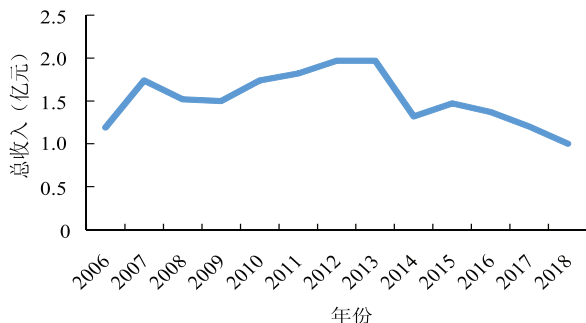


图2 2006—2018年大兴区休闲农业观光园经营总收入变化趋势

(2) 休闲农业的发展问题：大兴区休闲观光农业的发展，虽然初具规模，但是整体的旅游环境尚未成熟，仍存在如下问题：第一，品牌知名度不高、没有形成特色农业产业带。在受外部因素影响，部分农业观光园关、停、迁、转后，没有建立可以带动剩余观光园发展的区域特色品牌，从而导致近几年经营总收入一直呈下降趋势。第二，农业观光园经营方式单一、没有形成多元化发展。从收入构成上看，虽然大兴观光园的收入目前正向多种经营收入共同构成转变，但是仍然以门票收入、采摘收入和出售农产品收入为主。单一的果品采摘活动受天气和环境变化影响较大，不利于休闲观光农业长期稳定的发展。

2 都市农业发展面临的机遇

受城乡建设、北京大兴国际机场等项目开展的

影响,大兴区都市农业的发展面临农业生产规模连年缩减,农村青壮年劳动力资源不断向二、三产业转移等挑战。但是通过优化农业结构,加快农业功能转型升级,也能化挑战为机遇,促进大兴区都市农业的发展。

2.1 强化优势资源和传统活动

大兴区在发展过程中已经积累了很多优势资源,如榆垓镇的4A级景区北京大兴野生动物园,庞各庄镇的梨花节、西瓜节,魏善庄镇的月季节等等。通过综合优势资源,传承传统农业文化活动,大力发展观光农业、体验农业、创意农业,合理规划现代农业精品旅游线路,能够促进大兴区都市农业的发展。

2.2 引流北京大兴国际机场旅客

大兴国际机场作为华北的交通枢纽,地跨大兴区礼贤镇、榆垓镇,到2025年可满足旅客吞吐量7200万人次。随着该机场的建成并投入使用,便捷的配套交通、巨大的旅客量必将给大兴区都市农业的发展带来新的机遇。合理利用农业生态资源,将机场周边城镇规划为具有都市农业及文化旅游特色的小城镇。建成一批特色鲜明、功能完善的都市农业观光园,吸引游客促进都市农业的发展。

3 都市农业三产融合发展方向

在了解大兴区都市农业发展可以利用区域优势资源后,该文选取4A级景区北京大兴野生动物园的游客评论作为研究对象,通过构建LDA主题模型,来探索都市农业利用优势资源进行三产融合的发展方向。

3.1 LDA模型

LDA(Latent Dirichlet Analysis)模型在2003年由Blei等人首次提出,是一种非监督的机器学习技术^[7]。LDA属于一种生成模型,可以被看作是一个文档的产生过程。它由3层贝叶斯模型构成,包括文档、主题、词3个层面。相较于传统的文献计量学方法,主题模型能够深入挖掘文本词汇间的语义关系,在主题发现方面有着巨大的优势^[8]。

LDA模型认为每篇文档包含多个不同的主题,这些主题以一定的概率进行混合形成一篇文档,而每个主题又包含多个不同的词,这些词以一定概率进行混合形成主题。故在LDA模型中,一篇文档

生成的方式如下:①从狄利克雷分布 α 中取样生成文档*i*的主题分布 θ_i ;②从主题的多项式分布 θ_i 中取样生成文档*i*第*j*个词的主题 $z_{i,j}$;③从狄利克雷分布 β 中取样生成主题 $z_{i,j}$ 对应的词语分布 $\phi_{z_{i,j}}$;④从词语的多项分布 $\phi_{z_{i,j}}$ 中采样最终生成词语 $w_{i,j}$ 。

3.2 实验过程

该实验以北京大兴野生动物园的游客评论作为研究对象,利用爬虫技术,爬取了“美团”和“去哪儿网”两个电商平台上的9428条评论数据。在数据清洗部分,首先去掉重复评论;然后在完成中文分词的基础上,联合使用“百度停用词”、“哈工大停用词表”、“四川大学机器智能实验室停用词”和“中文停用词库”4个停用词表对评论文本去停用词;其次去掉分词后不足10个词的评论文本,剩余数据约6065条评论;最后将清洗后的评论数据按9:1的比例分为训练集和测试集。

在信息论中,困惑度(perplexity)用来度量一个概率分布或概率预测样本的好坏程度。低困惑度的概率模型分布或概率模型能更好地预测样本^[9]。计算公式为:

$$\text{perplexity} = e^{\frac{-\sum \log(p(w))}{N}} \quad (1)$$

$$p(w) = \sum p(z|d) * p(w|z) \quad (2)$$

其中, $p(w)$ 是指测试集中出现的每一个词的概率; N 表示测试集中出现的所有词,即测试集的总长度,不排重; z 表示训练过的主题; d 表示测试集中的各篇文档。

通过LDA模型分析文本数据时,需要先确定主题数量。按顺序选定主题数量,通过训练集训练LDA模型,再将测试集带入训练好的模型中计算此时的困惑度,实验结果如图3所示。根据困惑度越低概率模型能更好地预测样本的原则,选取主题数量为5,此时困惑度最小,说明选取的主题数量最合理。

3.3 实验结果与分析

确定主题数量为5后,对经过数据预处理的评论文本进行LDA主题模型建模。根据先验知识将参数设置为 $\alpha = 0.5$, $\beta = 0.1$,吉布斯抽样的迭代次数100次。

游客评论主题分类的实验结果如表1所示,每个主题分别选取前5个具有数据可分析性的主题词。由表1可知,游客评论主要涉及主体人群、出

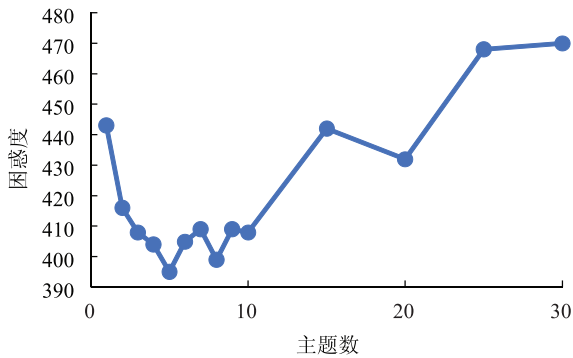


图3 大兴区野生动物园游客评论内容主题划分困惑度值

表1 大兴区野生动物园游客评论的主题分类

主题一	主题二	主题三	主题四	主题五
主体人群	出行方式	动物种类	互动形式	票务信息
孩子	自驾	老虎	胡萝卜	买票
小朋友	自驾游	狮子	孩子	购票
老人	自驾车	猛兽	食物	窗口
体验	服务	长颈鹿	蔬菜	价格
近距离	设施	大象	喂食	订票

行方式、动物种类、互动形式和票务信息五大方面。通过主体人群和活动形式的主题词可以分析得到，游客希望小朋友能在野生动物园里体验与动物、大自然的近距离互动，其中主要的互动方式是喂食动物。喂食环节中的食物来源涉及到农业中的所有产业，如第一产业中的农作物种养等，第二产业中的食品加工等，第三产业中的采摘旅游等，这些环节均可以设计成以“农业+教育科普”为主

题，适合亲子体验的休闲农业活动。大兴区可以深度开发农耕文化，将农业中的一、二、三产业融合发展，形成以北京大兴野生动物园的客流量为依托，“农业+教育科普”为主题的大兴区特色都市农业产业链。

4 结论与发展对策

大兴区都市农业在发展过程中，一方面受各种外界因素影响，粮食、蔬菜、花卉、瓜果等生产总规模连年减小，农业产值也逐年下降；另一方面，区域内休闲农业品牌知名度不高、农业观光园经营方式单一，没有形成特色农业产业带。这些问题在一定程度上阻碍了大兴区农业现代化的发展。但同时大兴区都市农业在发展过程中也积累了不少优势资源、遇到了北京大兴国际机场建成落地的新机遇。据此提出以下发展对策：（1）优化大兴区传统农业结构，提高种植效益，提升农产品的品质，做优、做精城市“菜篮子”项目来满足不同层次消费群体的生活。（2）深度开发农耕文化，依托区域优势资源、传统农业文化节、北京大兴国际机场等带来的客流量，以“农业+教育科普”、康养保健、农业文化传承等为主题，建成一批适合农业三产融合发展、主题特色鲜明、功能完善的新型都市农业观光园。（3）完善“大兴旅游”APP，在梨花节、西瓜节、月季节等传统活动期间，大兴旅游委可以规划合适的乡村旅游线路，引导游客延长游玩时间，增加游客到周边农业观光园、民俗旅游点体验农耕文化的可能性。

参考文献

- [1] 北京市大兴区农业农村局. 北京市大兴区“十三五”农业发展规划（2016-09-19） [2019-10-15], http://www.Beijing.gov.cn/zfxgk/dxq376/zzqgh/2016-09/19/content_dxq6486343.shtml.
- [2] 蒋和平, 卢江勇. 北京市发展都市型现代农业的基本思路与政策建议. 农业经济问题, 2007 (S1): 45-50.
- [3] 钟春艳, 王敬华, 文化. 北京都市型现代农业功能的哲学思考. 中国农业资源与区划, 2014, 35 (2): 47-52.
- [4] 孟素洁, 杨小琼. 北京都市型现代农业快速发展. 数据, 2009 (7): 11-12.
- [5] 北京市大兴区人民政府. 大兴区农业结构调整工作意见（2015-03-07） [2019-10-22], <http://www.bjdx.gov.cn/bjdxqrmzf/zwfw/zcwj/zfwj/633782/index.html>.
- [6] 江晶, 史亚军. 北京都市型现代农业发展的现状、问题及对策. 农业现代化研究, 2016, 36 (2): 168-173.
- [7] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation. Journal of Machine Learning Research, 2003 (3): 993-1022.
- [8] 王德川, 赵瑞雪, 寇远涛, 等. 基于LDA模型的欧盟框架计划近10年农业科研项目研究主题发现. 农业展望, 2017 (4): 69-75.
- [9] 刘雅姝, 张海涛, 徐海玲, 等. 多维特征融合的网络舆情突发事件演化话题图谱研究. 情报学报, 2019 (8): 798-806.

RESEARCH ON THE DEVELOPMENT STATUS OF URBAN AGRICULTURE IN DAXING DISTRICT BASED ON LDA MODEL

Pi Jingwen^{1,2}, Wang Haiyan^{1,2}, Wu Baoguo^{1,2}

((1. School of Information Science & Technology, Beijing Forestry University, Beijing 100083, China;

2. Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland
Administration, Beijing 100083, China)

Abstract In the sprint to fully realize agricultural modernization, the development of urban agriculture in Daxing district plays an important role in realizing regional agricultural modernization. The development status of urban agriculture in Daxing district was mainly analyzed from three aspects: facility agriculture, planting industry and leisure agriculture sightseeing park. The research on the integrated development of the primary, secondary and tertiary industries in agriculture mainly used the LDA topic model to analyze tourists' comments on the "Beijing Daxing wildlife park". After comprehensive analysis, a new development model of urban agriculture in Daxing district was proposed to improve the status quo of traditional urban agriculture in Daxing district.

Keywords urban agriculture; development status; LDA topic model

(上接第 38 页)

MATTERS NEEDING ATTENTION AND BUSINESS PROCESS OF PRELIMINARY EXAMINATION IN PATENT APPLICATION —TAKING CHINA (ZHONGGUANCUN) INTELLECTUAL PROPERTY PROTECTION CENTER AS AN EXAMPLE

Zhang Zhigang^{1,2}, Wang Wanxin^{1,2}, Chen Fang^{1,2}, Zhang Chaomin^{1,2}, Li Yansu^{1,2*}

(1. Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China;

2. Key Laboratory of Horticultural Crop Biology and Neoplasm Innovation, Ministry of Agriculture and Rural Affairs, Beijing 100081, China)

Abstract This paper briefly introduced the overall situation of China (Zhongguancun) intellectual property protection center and the patent classification of patent application pre-examination service, introduced the advantages, potential risks and business processes of patent application pre-examination in detail. And finally it provides reference for handling patent application pre-examination business.

Keywords patent; pre-examination; process