

Metadata Modeling and Empirical Research on the Description of Clothing Commodity Attributes in E-Commerce Platform

Li Zhou, Shixue Lu

The Glorious Sun School of Business and Management, Donghua University, Shanghai
Email: lucia_lsx@163.com

Received: Nov. 25th, 2019; accepted: Dec. 4th, 2019; published: Dec. 11th, 2019

Abstract

When conducting merchandise management on an e-commerce platform, a series of attributes are often used to describe the SKU. Typical attribute characteristics can distinguish each item well. There are many kinds of clothing product attributes, and the metadata under each attribute category is also very large and has a strong time-varying. However, attribute management for describing clothing goods lacks perfect metadata support, and problems such as incomplete product description and poor comprehensibility often appear. The paper takes the clothing products as an example, drawing on the Model Driven Architecture (MDA) proposed by OMG, building a metadata model for clothing products attribute management through Unified Modeling Language (UML). And this model is applied to the empirical study of the blouse commodity attribute data on an e-commerce platform. The constructed model is easy to expand and reuse, and provides theoretical support for e-commerce platform for commodity attribute data management.

Keywords

Clothing Goods, Attribute Description, Metadata Model, Empirical Research

电商平台服装商品属性描述元数据建模及实证研究

周力, 鲁仕雪

东华大学旭日工商管理学院, 上海
Email: lucia_lsx@163.com

收稿日期: 2019年11月25日; 录用日期: 2019年12月4日; 发布日期: 2019年12月11日

摘要

在电商平台进行商品管理时,通常运用一系列属性来描述SKU,典型属性特征能够很好地区分每个商品。服装商品属性种类繁多,各属性类别下的元数据表现体量也非常大,且有较强的时变性。而对服装商品属性描述缺乏完善的元数据支持,常常出现商品属性描述不完整、可理解性差等问题。论文以服装商品为例,借鉴OMG提出的元建模模型驱动架构(MDA),通过统一建模语言(UML)进行服装商品属性描述元数据建模。并将此建模方法应用于某电商平台女士衬衫商品属性数据的实证研究。构建的模型易于扩展、复用,为电商平台进行商品属性数据管理提供理论支撑。

关键词

服装商品, 属性描述, 元数据模型, 实证研究

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在电子商务蓬勃发展的背景下,天猫、京东、亚马逊等一系列电子商务平台纷纷涌现,吸引众多商家和消费者在这个平台上进行交易。服装类商品一直在电子商务销售中占据极其重要的地位,且已成为热门销售领域。当前,消费者在选择商品时已经关注到商品的具体属性,不仅积累了产品级别的体验,还积累了属性级别的体验[1] [2] [3]。在各大电子商务平台上,服装类商品的种类和数量也是越来越大,服装商品属性作为影响消费者感知的重要因素,对其的管理是目前电商研究领域值得重视的热点之一。

2. 研究现状

在信息管理[4] [5] [6]领域中,利用元数据建模,产生不同领域的应用元模型,提供系统的标准化建模方法和信息表达机制,可以有效提升信息管理的可维护性。同时,应用元模型,建立元模型和模型结构的约束也能利于信息管理中语言的扩展[7]。在利用元数据模型进行产品数据管理时,基于资源元建模的产品数据管理框架,将产品视为数据资源,可以识别出产品的特殊性[8]。形成专用于产品服务的通用元模型,还有助于处理支持决策过程的各种问题,以管理复杂系统(如 PSS) [9]。元建模技术用于面向 LCA 的过程表示时,可以用于产品信息的收集和组织,并确保数据的完整性、一致性和可追溯性[10]。构造需求的元模型,还能实现产品信息管理生命周期多目标优化[11]。

在进行电商平台应用和数据管理中,基于模型驱动架构(MDA)建立支持 Web 应用程序开发的电子商务元模型,确保了语义互操作性,同时可以降低复杂性并最大限度地减少创建和维护电子商务网站的工作量[12] [13] [14]。在 UML 上进行扩展,添加电子商务元模型及数据库建模结构,可以为 Web 开发和数据库应用实现元数据支撑框架[15]。采用统一建模语言 UML 和元建模语言 MOF,可以定义针对 web 应用的面向对象建模语言,使得系统更加易于扩展、复用[16]。当前的相关研究中更加侧重实现标准化的描述方法和通用的技术应用,很多模型追求通用性而缺少对不同商品信息的差异进行区分,另外,商品属性的种类多,不同属性的元数据表现也缺少具体的实证分析,从而缺乏商品管理元数据支持。

电商平台进行服装商品属性描述时,存在诸多问题[17] [18],如商品属性集合不全面,商品描述信息

杂乱, 极大地影响消费者的购物体验; 商家文化程度不同, 对商品结构理解不清晰, 常常按照自己的理解添加商品属性描述, 容易出现缺漏、错误标示等问题; 对于平台没有提供的属性类目, 商家还会通过商品标题来体现, 导致元数据的可理解性差, 数据管理成本提高。另外, 对服装类商品属性和商品描述的相关研究较少, 无论是理论还是知识运用都具有一定的局限性。周妍和王军[18]引用阮冈纳赞在图书馆学中的分面分类法, 以女装商品为例, 建立服装商品属性的分面元数据模型。中华人民共和国国家标准 GB/T 32670 [19]中制定了电子商务交易中服装类商品信息描述的方法。在 2019 年 7 月即将实施的国家标准 GB/T 37600.13 [20]规定了服装商品核心元数据的统一建模语言描述, 但是国家标准中提出的通用模型描述较笼统、泛化, 对不同属性的具体表现还需要另外分析, 缺乏较好的元数据支持, 无法用于信息化建设的指导。随着服装商品数据的日益增加, 需求的不断变动, 应用的规模和复杂度的日渐提高, 对于不同的商品应用传统的通用方法管理会难以维护、扩展。另外, 不同属性类别下的元数据表现具有一定的时代特征, 会随着时间的推进而变化, 对其的研究还需要立足于具体的实证数据分析上。

3. 电商平台服装商品属性描述元建模方法

对于电商平台的服装商品来说, 商品属性是构成整体产品的基本单位, 一件服装必须由一组属性组合而成。商家进行商品描述时运用的属性不仅要与商品本身的性质有关, 而且要能够与买家查找、挑选商品相联系。那就需要建立高效的元数据框架来对复杂多样的商品属性进行分类与管理。在论文中, 以服装商品作为研究对象, 进行服装商品属性描述元数据建模。

3.1. 元数据建模元层次

电商平台在进行商品信息及其属性的管理时, 对于不同的商品, 除了商品特殊属性的管理外, 对其他属性的管理要更加易于扩展、复用也是我们在进行服装商品属性管理值得重点考虑的。因此论文在建立电商平台服装商品属性描述的元数据模型时, 借鉴 OMG 提出的元建模模型驱动架构(MDA), 制定适用于特定领域的元模型, 从而大幅提升数据管理的效率, 构建的模型更加清晰的同时又易于后台的商品数据信息管理。建立电商平台服装商品属性信息管理的模型架构体系如图 1 所示。

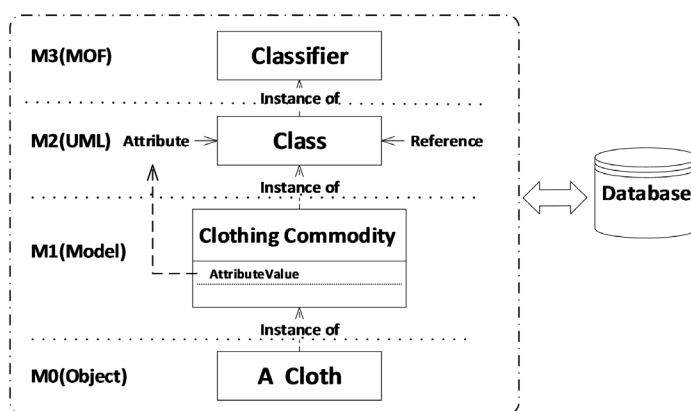


Figure 1. The meta-modeling architecture of clothing commodity attribute information management

图 1. 服装商品属性信息管理元建模架构

在对电商平台的服装商品属性信息进行管理时, 元模型层(M2)及模型层(M1)的数据管理是极为重要的。在元模型层(M2)中, 电子商品平台在进行商品数据管理时, 服装商品属性种类多, 数据繁杂等, 没有统一的标准, 需要提出一种规范化的元数据管理模型体系帮助平台进行数据管理。另外, 在模型层(M1)

中, 对于不同商品属性的表现还需要立足于具体商品的实证数据分析上, 能够挖掘出具体商品属性元数据模型区别于通用元数据模型的特殊性, 为电商平台商品数据管理及服装数据库等领域提供实证参考。

3.2. 元数据模型描述方法

在元数据模型中, 用 UML 类图体现商品属性实体和属性元素之间的关系, 用类的概念表示属性实体(商品属性), 用类的属性概念来表示属性元素(商品各属性类别下的元数据表现)。建立元数据模型时, 属性管理元数据模型描述如图 2 所示。将抽象的服装商品属性描述元数据模型表示为三元组 $C = (E, A, R)$ 。其中,

$E = (E_1, E_2, \dots, E_n)$ 是实体集合, 表示服装商品属性实体;

$A = (A_1, A_2, \dots)$ 是属性集合, 表示各属性实体下的属性元素(元数据表现);

$R = (R_p)$ 是语义关系集合, R_p 表示服装商品属性实体间的聚合关系。

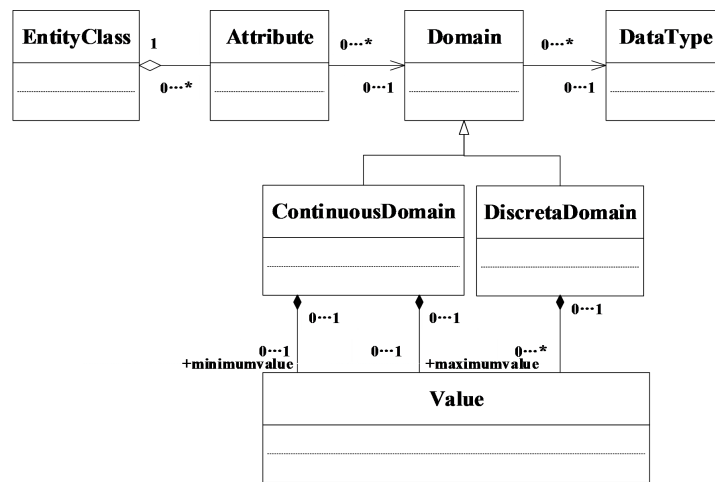


Figure 2. The metadata model description of attribute management

图 2. 属性管理元数据模型描述

4. 电商平台服装商品属性描述元数据模型

在分析元模型层的服装商品属性实体时, 通过比较当前国家标准 GB/T 37600.13 (2019.07.01 实施)中对服装商品属性核心元数据的统一描述中, 发现有以下几点不足:

1) 国家标准中提出的是一种通用的商品属性核心元数据统一描述, 没有对商品属性描述进行区分, 无法对商家进行商品描述提供指导和帮助。

2) 服装商品的基本属性有属性缺少的情况, 如服装的装饰细节、适用场合、穿着方式、组合形式等属性。

3) 商家在电商平台进行商品交易时, 相关营销及促销信息也是极其重要的。如 Lee H 和 Shin S 等[21]的文章中指出, 在浏览网站时, 消费者可能会遇到特别优惠, 不同颜色或理想的设计, 这可能反过来引发购买。Hong H 和 Lee SG [22]强调各种选择, 价格或促销以及感官属性是通过互联网鼓励服装购买意图的关键, 且研究的重点是与浏览相关的属性。因此, 在提出的模型中加入商品的营销属性。

4) 对于不同的商品, 除了有适用于所有商品的通用属性信息, 还有仅适用于该类商品的专有属性信息, 将通用属性及专有属性区分开可以更高效的对商品属性进行管理, 也帮助商家与消费者更清晰的区分不同的商品信息。

为了解决以上问题, 论文以电商平台服装商品为例, 提出电商平台服装商品属性描述的元数据模型如图3所示。

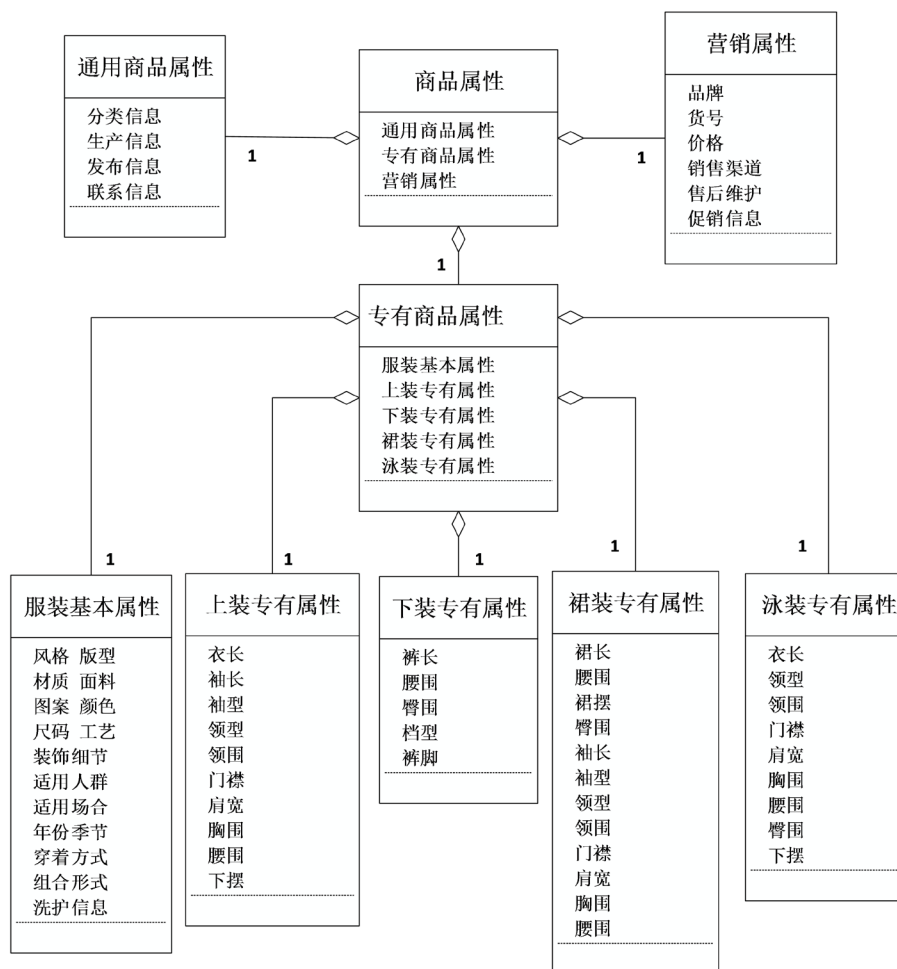


Figure 3. Metadata model of clothing commodity attribute in e-commerce platform
图3. 电商平台服装商品属性元数据模型

在模型层, 是对元模型层的实例化, 主要体现服装商品属性的具体元数据表现。当前没有针对服装商品属性元数据表现的相关规范化的元数据研究, 电商平台上的商家标示商品属性时, 常常出现属性乱用或者错误标示的情况, 缺少相关的元数据支持。以女士衬衫商品作为对象, 得出的模型层的元数据表现如图4所示(在图4中仅列出电商平台进行商品描述时的属性信息, 对于商品通用属性中的生产信息、发布信息、联系信息不在论文讨论范畴内)。在对模型层的元数据表现进行相关研究时, 还需立足于当前电商平台商品属性描述数据的实证分析上。论文将在第5章通过具体实证分析为电商平台服装商品属性描述提供元数据支持。

5. 电商平台服装商品属性运用实证研究

在上文中, 我们构建了一个有效的元数据模型来对繁杂多样的服装商品属性进行整序管理, 但是对模型层中具体元数据表现的描述还较简略。因此, 在该章节中, 我们对电商平台中女式衬衫商品属性描述数据进行实证研究, 了解属性的分布及应用情况, 进一步理解元数据模型中模型层的元数据表现情况。

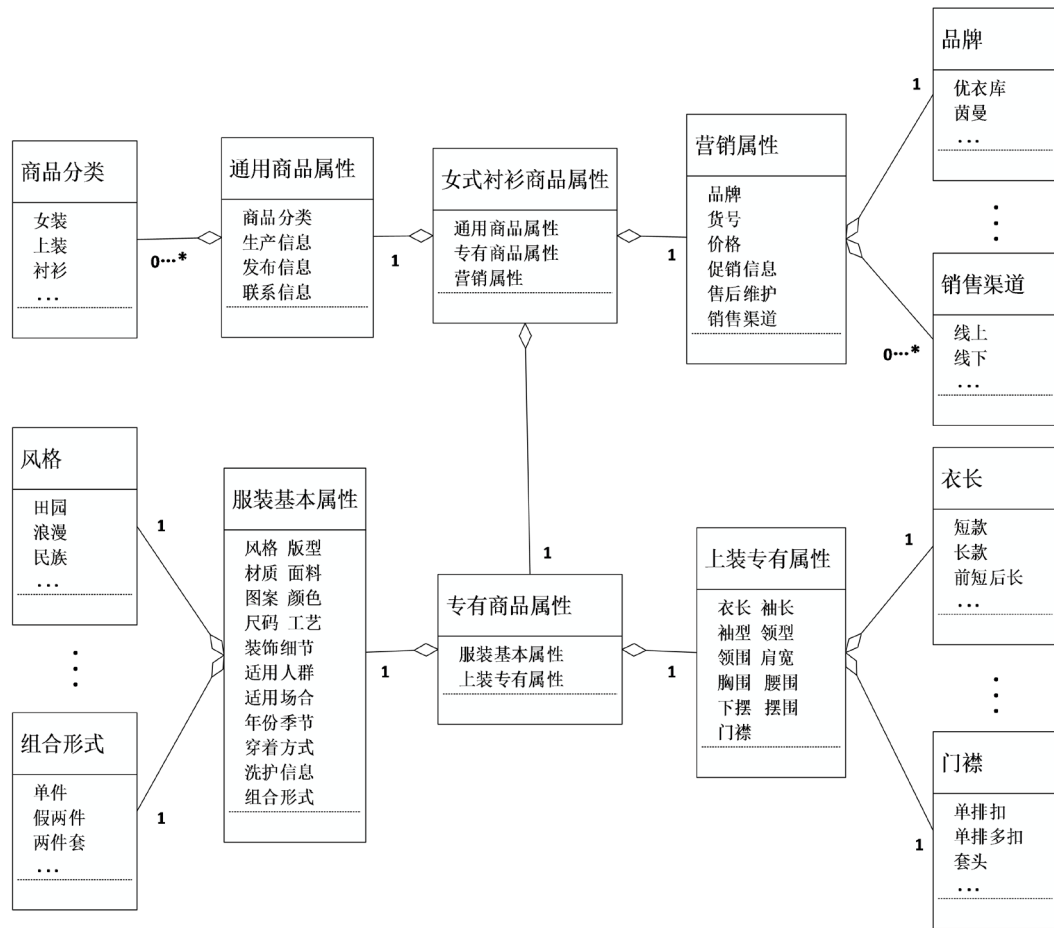


Figure 4. The metadata performance of blouse commodity in model layer
图 4. 女式衬衫商品模型层元数据表现

5.1. 商品属性分布

在当前的电子商务平台上, 服装商品属性情况通常以商品标题描述和商品详情情况来体现。通过比较天猫、京东、亚马逊三个平台对服装商品的描述发现, 天猫、京东两个平台的商品详情情况通常以结构化的方式呈现, 而亚马逊平台上的商品详情大多以非结构化的语言描述进行呈现。各平台上服装商品的详情中对服装商品属性的描述也不全面。

通过网络爬虫爬取天猫、京东两个平台在 2019 年 3 月月销量前 5000 的女士衬衫商品标题和商品详情情况, 经过数据整理后发现, 天猫平台服装商品详情中共出现了产品名称、品牌、货号、销售渠道类型、衣长、材质、面料、厚薄、风格、服装版型、服装款式细节、流行元素、服饰工艺、袖长、袖型、领型、衣门襟、颜色、图案、尺码、适用年龄、穿着方式、组合形式、年份季节、街头、甜美、通勤 27 个属性类别。京东平台服装商品详情中共出现了商品名称、商品编号、店铺、商品毛重、货号、腰型、风格、领型、使用年龄、图案、组合规格、主要材质、面料、廓型、流行元素、袖长、袖型、分类、衣门襟、上市时间、衣长、版型、商品产地、厚度、颜色、尺码、填充物、洗涤方式 28 个属性类别, 但是大部分商家在撰写商品详情时都仅列出 18~22 个左右的属性类别。其中, 天猫平台上出现属性类别乱用且有错误标示的情况, 如街头、甜美、通勤三个属性类别应该属于风格中, 另外对于服装适用场合、腰型等多个属性也未进行描述。而在京东平台上, 对服装风格的描述进行大量堆积, 无法识别商品的准确

风格, 另外, 对适用场合、服饰工艺等属性类别也未进行描述。

分析各个属性类别下的元数据表现后发现, 在服装商品属性描述中, 服装图案、服装装饰细节、风格、流行元素、服饰工艺、袖型、年份季节七个属性变化较大, 而其他属性较稳定。同时, 服装商品属性描述有以下几个特点: 1) 地域特点。不同区域的服饰由于当地民族文化特点的不同而有所区别, 如少数民族服饰。同时, 不同地区的气候特点也不尽相同, 如一些紫外线较强地区会凸显商品防晒功能, 较热区域凸显商品吸汗功能等。2) 时代特点。不同时间有不同的风格和流行元素。3) 季节性特点。在各个季节, 电商平台的服装商品有不同的特点, 如冬装凸显保暖性, 夏装凸显轻薄凉爽。

5.2. 商品标题属性运用现状

商品的属性有一定的时变性, 不同商品上架时间也不近相同, 因此很多商家通常在商品标题中加入当前比较重要的商品属性来吸引消费者购买, 另一方面消费者也更容易从搜索引擎中查找到该商品。在此次实证研究过程中, 爬取天猫平台上 2018 年 9 月、2018 年 11 月、2019 年 1 月和 2019 年 3 月四个月女士衬衫类商品销量前 2000 的商品标题信息, 分析商品标题中的属性运用情况。通过结巴分词统计商品标题中的商品属性, 以商品属性作为关键词, 运用 CiteSpace 对关键词进行聚类共现分析, 得出 2018 年 9 月、2018 年 11 月、2019 年 1 月和 2019 年 3 月四个月的商品标题中商品属性聚类共现网络图谱如图 5

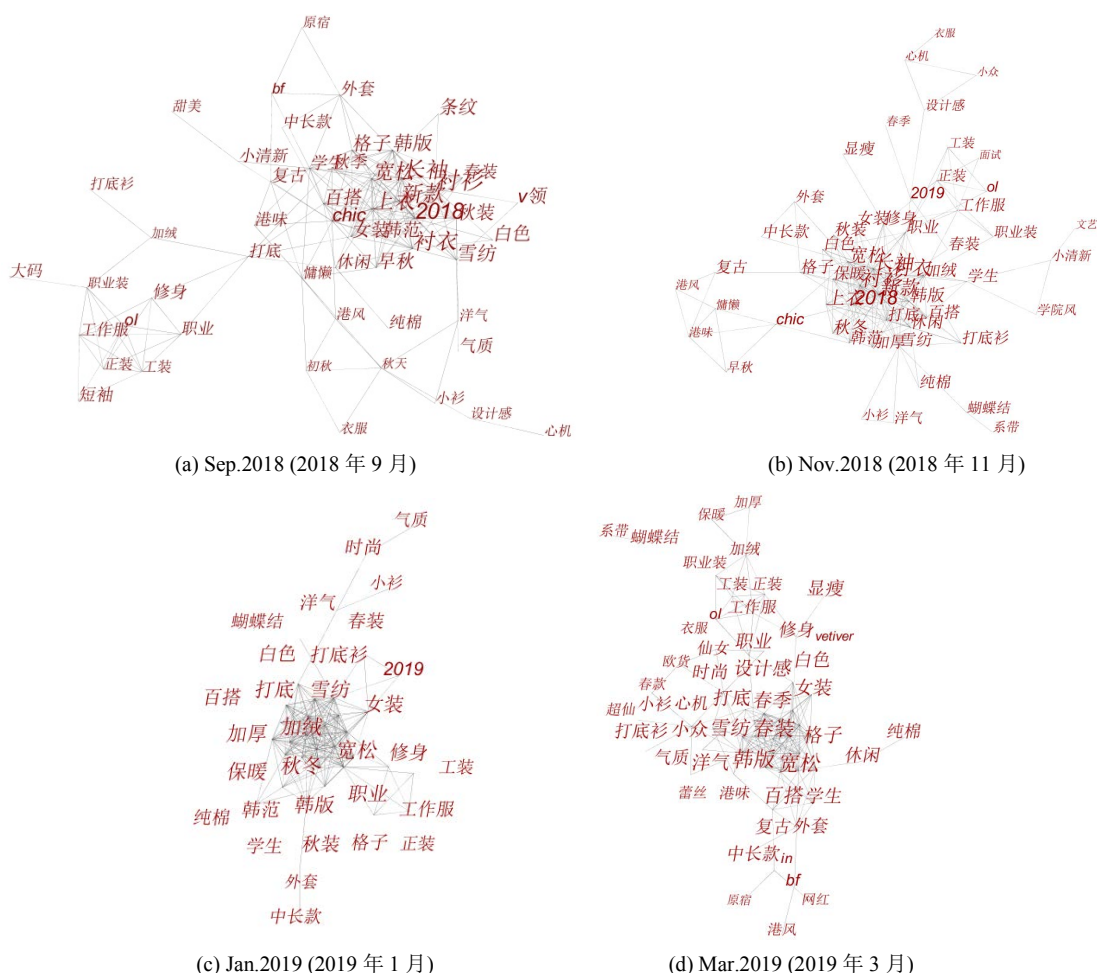


Figure 5. Attribute distribution clustering network diagram of blouse product title
图 5. 女式衬衫商品标题属性聚类共现网络图

所示。从图中可以发现商家在撰写商品标题时,常用的属性包括服装分类、年份季节、服装版型、风格、材质、袖长等属性。另外,商家在撰写商品标题时有很多组合特点。如“职业装”、“正装”类的商品标题中通常出现“工作服”、“工装”类的相似商品分类,还会出现“ol”的风格及“面试”、“商务”等工作场合的描述。商家撰写的“心机”这样的描述通常与“设计感”、“小众”、“漏肩”等属性描述出现。对于不同风格的商品都有其常用的属性,如“甜美”风格通常与“小清新”、“可爱”等描述同时出现。

6. 结论

论文借鉴元建模模型驱动架构(MDA)的思想,构建电商平台服装商品属性描述的元数据模型,使得电商平台进行商品属性数据管理时有较好的数据结构性且易于扩展。同时,构建的元数据模型为平台进行服装商品属性管理及服装数据库构建等领域提供元数据支持,利于信息化建设的指导。帮助商家进行商品管理时克服属性乱用、错误标示等问题。同时也帮助平台进行商品管理时克服商品属性描述非结构化、杂乱无章等问题,从而更高效地进行数据管理。最后通过实证数据分析挖掘出电商平台上服装商品属性分布和属性特征也为相关领域研究提供实证参考。

论文在进行电商平台服装商品相关属性研究中,仅考虑了电商平台进行商品描述的属性情况,未对商品发布信息、生产信息等进行研究,缺乏完整的元数据描述方案。另外,对于服装商品属性的具体表现具有较强的时变性,无法进行完整、详尽的总结。

参考文献

- [1] Ghani, R., Probst, K., et al. (2005) Text Mining for Product Attribute Extraction. *SIGKDD Explorations*, **8**, 41-48. <https://doi.org/10.1145/1147234.1147241>
- [2] Popescu, A.M. and Etzioni, O. (2005) Extracting Product Features and Opinions from Reviews. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, October 2005, 339-346. <https://doi.org/10.3115/1220575.1220618>
- [3] Park, E.J., Kim, E.Y., et al. (2011) Apparel Product Attributes, Web Browsing, and e-Impulse Buying on Shopping Websites. *Journal of Business Research*, **65**, 1583-1589.
- [4] Chantilly, R.K.S. (2015) UMLS Semantic Network as a UML Meta-Model for Improving Biomedical Ontology and Application Modeling. *International Journal of Healthcare Information Systems and Informatics*, **10**, 34-56. <https://doi.org/10.4018/IJHISI.2015040103>
- [5] Yang, B., Qiao, L.H., Cai, N., et al. (2018) Manufacturing Process Information Modeling Using a Meta-Modeling Approach. *The International Journal of Advanced Manufacturing Technology*, **94**, 1579-1596. <https://doi.org/10.1007/s00170-016-9979-0>
- [6] Yu, L., Wang, S.Y. and Lai, K.K. (2009) A Neural-Network-Based Nonlinear Metamodeling Approach to Financial Time Series Forecasting. *Applied Soft Computing*, **9**, 563-574. <https://doi.org/10.1016/j.asoc.2008.08.001>
- [7] Alanen, M. and Porres, I. (2008) A Metamodeling Language Supporting Subset and Union Properties. *Software and System Modeling*, **7**, 103-124. <https://doi.org/10.1007/s10270-007-0049-9>
- [8] Qin, S.J., Cai, H.M. and Jiang, L.H. (2012) A Product Life Cycle Data Management Framework Based on Resource Meta-Modeling. *IEEE Asia-Pacific Services Computing Conference*, Guilin, 6-8 December 2012, 305-308.
- [9] Boucher, X. and Medini, K. (2016) Towards a Generic Meta-Model for PSS Scenarios Modelling and Analysis. *Procedia CIRP*, **44**, 234-239. <https://doi.org/10.1016/j.procir.2016.03.038>
- [10] Tao, J. and Yu, S.R. (2018) A Meta-Model Based Approach for LCA-Oriented Product Data Management. *25th CIRP Life Cycle Engineering Conference*, Copenhagen, 30 April-2 May 2018, 423-428. <https://doi.org/10.1016/j.procir.2017.12.010>
- [11] Carreras, J., Pozo, C., Boer, D., et al. (2016) Systematic Approach for the Life Cycle Multi-Objective Optimization of Buildings Combining Objective Reduction and Surrogate Modeling. *Energy and Buildings*, **130**, 506-518. <https://doi.org/10.1016/j.enbuild.2016.07.062>
- [12] Tarazona, G.M., Pelayo, B.C., Sanjuán, O., et al. (2014) e-Commerce Metamodel Based on MDA. *IEEE Latin Ameri-*

-
- ca Transactions*, **12**, 699-705. <https://doi.org/10.1109/TLA.2014.6868872>
- [13] Huang, S.-M., Hua, J.-S., Will, H., *et al.* (2012) Metamodeling to Control and Audit E-Commerce Web Applications. *International Journal of Electronic Commerce*, **17**, 83-118. <https://doi.org/10.2753/JEC1086-4415170104>
- [14] Figueiredo Adriana Maria, C.M. and Aqueo, K. (2004) Using Metamodels to Promote Data Integration in an e-Government Application Scenario. *IFIP Advances in Information and Communication Technology*, **139**, 293-303. https://doi.org/10.1007/1-4020-7907-9_23
- [15] 张芳, 史殿习, 罗端红. 元建模在电子商务开发平台中的研究与应用[J]. 计算机应用研究, 2005, 22(9): 191-193, 213.
- [16] 刘家红, 吴泉源. Web 应用开发平台的元建模研究[J]. 计算机工程与科学, 2006, 28(3): 88-91.
- [17] 王军, 周妍. 电商平台的分面元数据研究[J]. 情报杂志, 2016, 35(3): 317-325.
- [18] 周妍, 王军. 电商平台商品信息组织机制研究[J]. 图书馆建设, 2016(6): 21-26.
- [19] 中国国家标准化管理委员会. GB/T 32670. 电子商务交易产品描述(服装)[S]. 北京: 中国标准出版社, 2016.
- [20] 中国国家标准化管理委员会. GB/T 37600.13. 全国主要产品分类-产品类别核心元数据[S]. 北京: 中国标准出版社, 2018.
- [21] Lee, H., Shin, S. and Kim, S. (2010) Surrogate Internet Shopping Malls: The Effects of Consumers Perceived Risk and Product Evaluations on Country-of-Buying-Origin Image. *Journal of Global Academy of Marketing Science*, **20**, 208-218. <https://doi.org/10.1080/12297119.2010.9730193>
- [22] Hong, H. and Lee, S.G. (2005) Effects of Apparel Merchandise on Experienced Emotion for Apparel Shopping and Attitude toward the Internet Store. *Korean Society of Clothing and Textiles*, **29**, 478-490.