

小说人物的分布表示及其应用研究*

贾玉祥^{1,2}, 王璐¹, 刘鹏程¹, 王钤³, 张岳⁴, 咎红英¹

(1.郑州大学 信息工程学院, 河南郑州 450001; 2.闽江学院 物联网产业化与智能生产协同创新中心, 福建福州 350108; 3.郑州大学 外语学院, 河南郑州 450001; 4.西湖大学 工学院, 浙江杭州 310024)

摘要: 小说是以刻画人物为中心, 通过完整的情节和具体的环境描写反映社会生活的一种文学体裁。对小说人物进行建模, 是小说文本理解和小说文本挖掘的基础性工作。本文构建了大规模的小说语料库, 抽取人物及其依存特征, 提出基于 skip-gram 的人物向量训练方法, 以人物为目标, 以依存特征为上下文。基于训练出的人物向量, 探索了小说人物相似度计算、小说人物聚类分析及小说人物画像等应用。实验结果表明, 小说人物的分布表示有较好的应用效果。

关键词: 小说人物, 分布表示, 人物向量, 人物聚类, 人物画像

Distributed Representation of Fictional Characters and Its Applications

JIA Yuxiang^{1,2}, WANG Lu¹, LIU Pengcheng¹, WANG Qian³, ZHANG Yue⁴, ZAN Hongying¹

(1.School of Information Engineering, Zhengzhou University, Zhengzhou, 450001, China; 2.Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University, Fuzhou, 350108, China; 3.School of Foreign Languages, Zhengzhou University, Zhengzhou, 450001, China; 4.School of Engineering, Westlake University, Hangzhou, 310024, China)

Abstract: Novel is a literary genre that centers on character creation, and reflects social life through complete plots and specific environmental descriptions. Modeling fictional characters is the basic work for literary text understanding and literary text mining. Firstly, we construct a larger-scale novel corpus and extract characters and their dependency features. Then we propose a skip-gram-based model to train character embeddings, the character as the target while the dependency features as the contexts. Based on the trained character embeddings, the applications of character similarity computation, character clustering, and character profiling are explored. The experimental results show a good performance of the distributed representation of fictional characters in the above applications.

Keywords: fictional character, distributed representation, character embedding, character clustering, character profiling

1 引言

计算文学^[1]是社会科学和信息科学的交叉学科, 研究内容包括文学作品自动分析^[2]、文学作品自动生成^[3]、计算文体学^[4]等。自然语言处理技术在计算文学中扮演着重要的角色。作为主要的文学体裁, 小说以刻画人物为中心, 通过完整的情节和具体的环境描写反映社会生活。面向小说人物^[5-7]的自然语言处理任务包括小说人物的表示及自动分类、人物关系的表示及自动分类、人物关系网络的抽取及定量分析等, 这些工作不仅可以推动文学研究的进展, 也可以为小说推荐、问答、翻译等商业应用提供算法基础。

深度学习技术在自然语言处理的很多任务中已经得到了广泛的应用, 然而在文学领域文本中的应用还很少。本文研究小说人物建模这一计算文学的基本问题, 从基础的分布式表示方法入手, 探索深度学习方法在文学领域文本中的应用。我们构建了大规模的小说语料库, 识别并抽取人物及其依存特征, 借鉴词向量的思想训练人物向量, 在此基础上进行了人物相似度计算、人物聚类

*收稿日期:

定稿日期:

基金项目: 国家自然科学基金 (61402419); 国家社会科学基金 (14BYY096); 闽江学院物联网产业化与智能生产协同创新中心开放基金 (IIC1707); 河南省科技攻关项目 (192102210260)

作者简介: 贾玉祥 (1981—), 男, 讲师, 主要研究方向为自然语言处理; 王璐 (1994—), 女, 硕士, 主要研究方向为自然语言处理; 刘鹏程 (1997—), 男, 硕士生, 主要研究方向为自然语言处理。

分析及人物自动分类的研究。

本文的章节安排如下：第2节介绍相关研究工作，第3节介绍小说语料库的构建、依存特征的抽取及人物的分布式表示方法，第4节展示人物相似度计算及实验结果，第5节展示人物聚类分析及实验结果，第6节展示人物画像及实验结果，最后给出总结与展望。

2 相关研究

人物建模是由文学作品文本自动生成人物的表示。Zhang et al.^[8]从故事文本中抽取人物的显著特征，形成人物的描述，进而生成故事摘要。Bamman et al.^[9]使用主题模型从电影情节摘要中学习人物类型，每个人物类型按依存关系从 agent、patient、attribute 三个方面进行表示，每一个方面表示为一系列隐含主题分布，每一个隐含主题又表示为词语的分布。Bamman et al.^[10]使用主题模型从小说文本中学习人物类型，每个人物类型按依存关系从 agent、patient、possessive、predictive 四个方面进行表示，每一个方面直接表示为词语的分布，在生成词语时加入了小说作者因素。

人物画像 (character profiling) 是一种显式的人物建模方式，由文学作品文本来预测人物的属性，如性别、年龄、性格等，是一个自动分类问题。Flekova and Gurevych^[11]预测小说人物性格是内向还是外向，使用 SVM 分类器，分别使用人物说的话、人物的动作、人物的形容词或副词性描述三种特征，加入了多种词汇资源，如 WordNet、VerbNet、LIWC、词向量等，结果显示使用人物说的话作为特征效果不如后两种特征。

人物关系自动分类。Massey et al.^[12]把小说人物关系粗分为社会关系 Social、工作关系 Professional 和家庭关系 Familial 三种，每一种关系再进一步细分；按关系的紧密程度分为友好 Positive、一般 Neutral 和对立 Negative 三种，并为 109 部英文小说标注了人物关系。借鉴“敌人的敌人是朋友”这种思想，Srivastava et al.^[13]在基于电影情节摘要来预测人物之间是友好还是对立关系时，引入了文本特征之外的结构特征，即目标人物与其他人物之间的关系，这对判断目标人物之间的关系是有用的。

人物关系变化建模。小说的篇幅较长，随着情节的发展，人物之间的关系往往会发生变化，这时不必显式定义关系类型，可以采用词语的分布来表示关系。关系的变化是一个序列问题，Iyyer et al.^[14]提出基于循环神经网络 RNN 和词典学习的人物关系模型 (Relationship Modeling Network, RMN)，关系类型用向量表示，用与关系类型向量最相似的一组词语来描述关系类型。Chaturvedi et al.^[15]使用带高斯发射的隐马尔科夫模型 (Hidden Markov Model with Gaussian Emissions, GHMM) 对人物关系变化进行建模，考虑了依存动词、词袋、语义框架等多种特征。Fremmann and Szarvas^[16]扩展了 RMN，同时对人物变化和人物关系变化进行建模，两种变化综合起来对小说进行表示，计算小说之间的相似度，进而实现小说推荐。

人物关系网络抽取及分析。抽取小说中的人物及关系，构造人物关系网络，通过研究网络的性质来发现或验证小说人物及其所处社会环境的特点。网络的节点是人物，边是人物之间的关系。关系的定义可以是人物之间的共现关系^[17]，即人物共同出现在一定尺度的上下文中，如一句话、一个段落或者一个章节；也可以是事件关系，即人物共同参与同一个事件^[18]；也可以是对话关系，即两个人之间有对话交互^[19]；也可以是某些显式定义的关系，如家庭关系、工作关系等。

以上研究工作的基础是文学领域文本中的分词^[20]、命名实体识别^[21]、对话者识别^[22]等基本任务，目前这些任务在新闻领域文本上的表现较好，而在文学领域文本中则面临着领域迁移带来的巨大挑战，逐渐引起了更多研究者的关注。

3 基于依存特征的小说人物表示

3.1 依存特征抽取

由于能够比较便捷地获取到大规模的过了版权期的英文小说，且英文文本的分析技术相对成熟，本文的研究从英文小说入手。从 Gutenberg Project¹及其他网站下载免费英文图书 24990 本，

¹ <http://www.gutenberg.org/>

使用 BookNLP^[10]进行词法分析、命名实体识别、依存句法分析、人名聚类、指代消解等处理，抽取与每一个人物存在特定依存关系的词及关系类型作为特征。依存关系选择 *nsubj*/主语、*dobj*/宾语、*nsubjpass*/被动主语、*poss*/所有格及 *amod*/形容词修饰五种类型^[23]，依存特征是指目标人物作为主语、宾语、被动主语的那些动作、目标人物拥有的事物、修饰目标人物的修饰语。

对人物与依存特征出现的频次加以限制，要求人物出现 10 次以上、特征出现 10 次以上，最后得到 436761 个人物及 44919 个特征。依存特征按类型分布情况如表 1 所示，第 2 列指的是该类型特征的词形 (*type*) 个数。可见，*poss* 与 *nsubj* 类型的特征词形分布较广。表 1 也给出了各类型特征中出现次数最多的 10 个词，其中 *poss* 类型最常见的是人所拥有的外貌 (*hand/eye/face/head*)、社会关系 (*father/mother/friend*) 及其他抽象状态 (*life/mind/heart*)。

表 1 依存特征分布情况

依存类型	个数	示例 top10
<i>nsubj</i>	14162	say, have, be, go, come, see, know, take, make, do
<i>dobj</i>	3881	tell, ask, see, take, cry, leave, meet, know, find, reply
<i>nsubjpass</i>	2673	go, call, bear, take, make, oblige, send, leave, know, get
<i>poss</i>	19339	hand, eye, face, head, father, mother, friend, life, mind, heart
<i>amod</i>	4864	old, poor, little, answer, young, major, good, great, exclaim, last

3.2 人物向量及特征向量学习

原始的 *skip-gram* 模型^[24]，由目标词 w_t 预测一定窗口内的上下文词，如 $w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}$ 。Levy and Goldberg^[25]不使用窗口选择上下文，而是把与目标词有句法依存关系的词及依存关系类型作为上下文。本文把人物作为目标词，把人物的五类依存特征作为上下文，采用 *skip-gram* 模型及负采样方法，同时训练出人物向量及依存特征向量。

INPUT PROJECTION OUTPUT

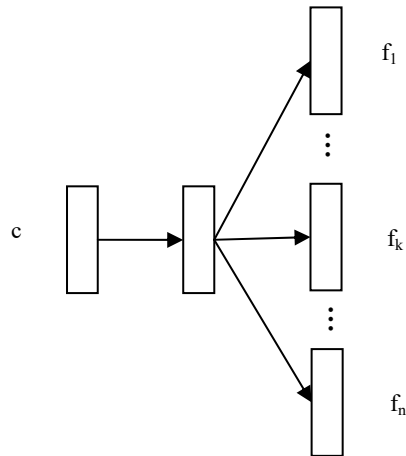


图 1 人物表示的 *skip-gram* 模型

如图 1 所示网络，输入为人物 c ，输出为其依存特征 f_1, f_2, \dots, f_n 。对于不同的 c ，依存特征个数 n 不尽相同。给定数据集 D ， (c, f) 是人物-依存特征对，模型最大化公式 1。

$$\prod_{(c,f) \in D} p(f|c) \quad (\text{公式 1})$$

对于负采样方法，最大化的目标函数如公式 2 所示， $(c, f) \in D$ 为正例， $(c, f) \in D'$ 为负例， $\sigma(\cdot)$ 为 sigmoid 函数， v_c 为人物 c 对应的向量， v_f 为依存特征 f 对应的向量。

$$\sum_{(c,f) \in D} \log \sigma(v_c \cdot v_f) + \sum_{(c,f) \in D'} \log \sigma(-v_c \cdot v_f) \quad (\text{公式 2})$$

模型的实现基于 word2vec^[25]，输入为人物和依存特征对，人物向量和依存特征向量的维度设为 200，负样本个数设为 10，迭代次数设为 20。

4 人物相似度计算

如问题 Q1 与 Q2 所示，网上存在大量询问相似人物的问题，这类问题的解答可以归结为人物相似度计算。人物相似度计算也是人物聚类分析与自动分类的基础。

Q1: Who is the other fictitious female character like **Elizabeth Bennet**?

Q2: Is there any other fictional character like **Sherlock Holmes**?

训练出人物向量之后，可以计算任意两个人物之间的相似度，相似度使用人物向量之间的夹角余弦来表示。表 2 和表 3 分别给出小说 *Pride and Prejudice* 中的 Miss Elizabeth Bennet 和小说 *The Adventures of Sherlock Holmes* 中的 Mr. Sherlock Holmes 两个人物最相似的 10 个人物，人物用“作者@作品名@人物名”来表示。

由表 2 可见，一个作者会在不同作品中塑造相似的人物形象，10 个人物中有 6 个都来自同一个作者 Jane Austen 的作品。还有 3 个人物来自 Maria Edgeworth 的作品，该作者在今天没有像 Jane Austen 一样为人熟知。通过查资料发现，两个作者确实存在某种关联，她们在当时都很欣赏对方的作品，都受到对方作品的影响。目前也有很多研究工作是比这这两位作者及其作品的。

由表 3 可见，我们计算的相似人物能够把同一个人物在系列作品中出现或被其他作品提及的情况找出来，这些都应该指的是同一个人物，理应具有较高的相似度。10 个人物中有 7 个都是 Mr. Sherlock Holmes，其他几个人物也是由别的作者塑造的侦探角色。由此可见，基于人物向量计算的相似人物可以回答 Q1 和 Q2 这类问题，可以为自动问答和小说人物推荐提供支持。

表 2 Austen,Jane,1775-1817@PrideandPrejudice@MissElizabethBennet

人物	相似度
Austen,Jane,1775-1817@MansfieldPark@FannyPrice	0.758
Austen,Jane,1775-1817@Emma@EmmaWoodhouse	0.743
Edgeworth,Maria,1767-1849@TalesandNovels—Volume07@MISSCAROLINEPERCY	0.733
Austen,Jane,1775-1817@Persuasion@MissAnneElliot	0.729
Edgeworth,Maria,1767-1849@TalesandNovels—Volume10@Helen	0.728
Austen,Jane,1775-1817@SenseandSensibility@Elinor	0.718
Edgeworth,Maria,1767-1849@TalesandNovels—Volume10@LadyCeciliaDavenant	0.704
Austen,Jane,1775-1817@MansfieldPark@MissCrawford	0.694
Burney,Fanny,1752-1840@Cecilia;Or,MemoirsofanHeiress—Volume2@CeciliaBeverley	0.690
Austen,Jane,1775-1817@NorthangerAbbey@CatherineMorland	0.672

表 3 Doyle,ArthurConan,Sir,1859-1930@TheAdventuresofSherlockHolmes@Mr.SherlockHolmes

人物	相似度
NoAuthor@TheLockandKeyLibraryClassicMysteryandDetectiveStories:ModernEnglish@Mr.SherlockHolmes	0.747
NoAuthor@TheBoyScoutsBookofStories@Mr.SherlockHolmes	0.644
NoAuthor@MasterpiecesofMysteryInFourVolumesDetectiveStories@Mr.SherlockHolmes	0.635
Various@ShortStoriesofVariousTypes@Mr.SherlockHolmes	0.608
Doyle,ArthurConan,Sir,1859-1930@TheReturnofSherlockHolmes@Mr.SherlockHolmes	0.578
Doyle,ArthurConan,Sir,1859-1930@TheMemoirsofSherlockHolmes@Mr.SherlockHolmes	0.560
Finnemore,John,1863-1915@JackHaydon'sQuest@Mr.Buxton	0.513
Reeve,ArthurB.(ArthurBenjamin),1880-1936@TheSilentBullet@CRAIGKENNEDY	0.506
Doyle,ArthurConan,Sir,1859-1930@TheSignoftheFour@Mr.SherlockHolmes	0.501
Doyle,ArthurConan,Sir,1859-1930@MicahClarkeHisStatementasmadetohisthreegrandchildrenJoseph,GervasandReubenDuringtheHardWinterof1734@DecimusSaxon	0.501

5 人物聚类分析

基于训练出的人物向量可以对人物进行聚类分析，聚类算法采用 **k-means**，评价指标采用纯度 **Purity**^[26]。如公式 3 和 4 所示，一个聚簇（cluster）的纯度用该聚簇中样本最多的类的样本占比来表示，整个聚类的纯度用各个聚簇纯度按聚簇大小加权求和得到。

$$\text{Purity}(C_r) = \frac{1}{n_r} \max_i(n_r^i) \quad (\text{公式 3})$$

$$\text{Purity} = \sum_{r=1}^k \frac{n_r}{n} \text{Purity}(C_r) \quad (\text{公式 4})$$

人物聚簇中作者的分布情况如表 4 所示。设计两组实验：在第一组中，每个人物出现次数不低于 100 次，每个作者塑造人物数不少于 50 个，最终得到 268 个作者塑造的 34869 个人物，聚簇个数 k 设为 268，结果 **Purity** 等于 0.333，即平均每一个聚簇中属于同一作者的人物约占 1/3。第二组中，每个人物出现次数不低于 200 次，每个作者塑造人物数不少于 50 个，最终得到 130 个作者塑造的 14171 个人物，聚簇个数 k 设为 130，结果 **Purity** 等于 0.473，即平均每一个聚簇中属于同一个作者的人物占比接近 1/2。这也反映了同一个作者塑造的人物具有相似性这一点。

表 4 聚簇中作者的分布情况

人物出现次数	每一作者塑造人物数	作者个数	人物个数	Purity
≥ 100	≥ 50	268	34869	0.333
≥ 200	≥ 50	130	14171	0.473

进一步观察一个聚簇中的作者分布，发现不同作者也会塑造类似的人物形象，因此可以发现作者之间的相似性或者关联。图 2 给出了编号为 52 的聚簇中作者分布的云图，作者出现次数与字体大小成正比，可见包含了大量 **Jane Austen** 小说中的人物和 **Maria Edgeworth** 小说中的人物，这也反映了两位作家之间的关联。类似地，图 3 给出的编号为 59 的聚簇中的作家 **Harry Collingwood** 与 **James Fenimore Cooper** 也塑造了大量类似的人物，原因是二者都写海洋探险小说或海盗小说，在 *The Jolly Roger Tales: 60+ Pirate Novels, Treasure-Hunt Tales & Sea Adventures* 这本书中列出了二者的相关作品。

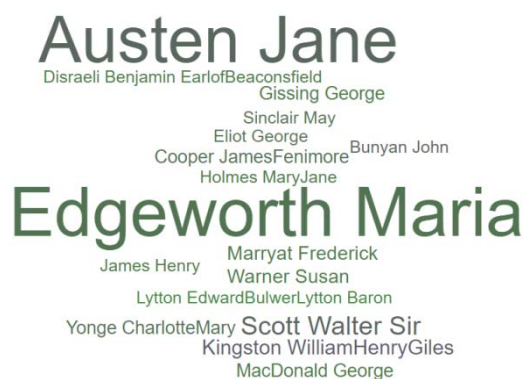


图 2 cluster-52 中作者分布



图 3 cluster-59 中作者分布

6 人物画像

人物画像 是根据描写人物的文本来预测人物的属性，包括人物的性别、年龄、性格等。这里基于人物参与的事件（`nsubj/dobj/nsubjpass`）、所有物（`poss`）及修饰语（`amod`）对小说人物的性别和婚姻状况进行预测，性别分为男女，婚姻状况分为已婚和未婚，因此是一个二分类问题。分类器采用多层感知机（Multilayer Perceptron, MLP）模型，输入向量维度为 200，隐藏层数为 1 层，隐藏节点数为 150，隐藏节点激活函数为 ReLU，输出节点激活函数为 Sigmoid，训练迭代次数为 100，学习率设为 0.05。

根据人物表示方法不同，输入向量采用三种不同的表示方式：人物向量 CE、特征向量平均 AFE 以及词向量平均 AWE。人物向量和特征向量的训练如第 3 节所述，AFE 指用人物对应的所有依存特征的向量的均值表示该人物，作为 MLP 的输入。AWE 是用人物对应的依存特征词的词向量的均值表示该人物，这里采用 GloVe^[27]预训练的 200 维词向量。

由于对人名进行了聚类，根据称谓中的 Mr.、Miss、Mrs. 很容易判断人物的性别，Mr. 对应男性，Miss 和 Mrs. 对应女性；婚姻状况方面，Miss 对应未婚，Mrs. 对应已婚。限制人物出现次数为 200 次以上，得到男性人物 4506 个，女性人物 3801 个，未婚人物 2204 个，已婚人物 1597 个。根据多数原则得到一个分类准确率 Accuracy 的基线，性别分类为 0.542，婚姻状况分类为 0.580。

表 5 人物画像实验结果

	CE	AFE	AWE
性别	0.991 ± 0.002	0.994 ± 0.002	0.907 ± 0.007
婚否	0.840 ± 0.010	0.838 ± 0.009	0.694 ± 0.012

采用五折交叉验证，分类实验的结果如表 5 所示（平均分类准确率±标准差）。可见，人物向量 CE 和特征向量平均 AFE 两种人物表示方式的分类结果类似，好于通用词向量平均 AWE 的表示方法，也体现了小说语料训练出来的向量在小说人物分类效果上优于通用词向量。三种人物表示方法的分类准确率都高于基线准确率。从两个任务分类效果的比较来看，性别预测已经达到 99% 的准确率，婚姻状况的预测要难的多。对于人物向量中没有出现过的人物，可以采用特征向量平均这种表示方法，能获得与人物向量相当的分分类效果，只要人物有特征具有向量表示即可。

可以借助词云进一步观察 Mr.、Miss、Mrs. 三类人物的依存特征。按照特征的相对频率（在该类人物中出现的频次除以在所有人物中出现的频次）降序排列，选择特征频次在 1000 以上的前 30 个特征构造词云，字体大小与特征在该类人物中出现的频次成正比，如图 4~图 6 所示。可见，Mr. 的显著特征有 `wife@poss`、`man@nsubj`、`gentleman@nsubj` 等；Miss 的显著特征有 `eye@poss`、`lip@poss`、`cheek@poss` 等与外表有关的词语，`father@poss`、`mother@poss` 等表示长辈关系的词语，`girl@nsubj`、`woman@nsubj` 等表示性别的词语等；Mrs. 的显著特征有 `husband@poss`、`daughter@poss`、

son@poss、child@poss 等表示丈夫、子女关系的词语及 woman@nsubj 等表示性别的词语。三类人物的显著特征既具有独特性，也有重叠，比如 Miss 与 Mrs.的共有特征 woman@nsubj，这也是性别容易区分，而婚姻状况（Miss 与 Mrs.）较难区分的一个原因。

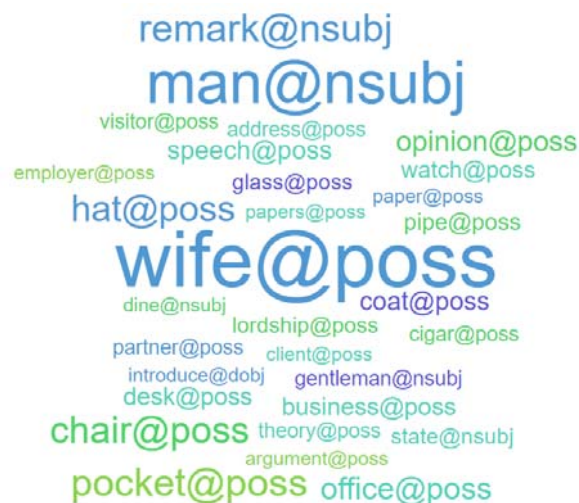


图 4 Mr.显著特征



图 5 Miss 显著特征



图 6 Mrs.显著特征

7 总结与展望

本文提出一种小说人物的分布式表示方法，基于人物依存特征和 skip-gram 模型学习人物向量，将学习出的人物向量应用于人物相似度计算、人物聚类分析及人物画像三个任务，表明了人物向量具有较好的表示效果。与传统的高维稀疏向量表示相比，在存储空间和运算时间上也具有巨大的优势。

通过人物相似度计算，可以考察不同作品中的相似人物，支持小说人物相关的自动问答。通过人物聚类分析，发现同一个作者在不同作品中塑造的人物形象往往具有相似性，也能发现塑造类似人物形象的不同作者之间的关联。人物画像的实验表明，通过小说文本中对人物的描述来预测人物属性是可行的。

下一步我们将使用循环神经网络对小说人物进行建模，考虑人物随情节发展的变化因素，探索人物的变化轨迹。对人物类型进行界定，实现显式的人物自动分类。构建中文小说语料库，研究中文小说人物的计算问题，并进行跨语言的对比研究，也是我们下一步的目标。

参考文献

- [1] 俞士汶, 王治敏, 朱学锋. 文学语言与自然语言理解研究[C]//中国中文信息学会. 中文信息处理前沿进展——中国中文信息学会二十五周年学术会议论文集. 2006: 72-79.
- [2] Hammond A, Brooke J, Hirst G. A tale of two cultures: Bringing literary analysis and computational linguistics together[C]//Proceedings of the Workshop on Computational Linguistics for Literature. 2013: 1-8.
- [3] Fan A, Lewis M, Dauphin Y. Hierarchical neural story generation[J]. arXiv preprint arXiv:1805.04833, 2018.
- [4] Neal T, Sundararajan K, Fatima A, et al. Surveying stylometry techniques and applications[J]. ACM Computing Surveys (CSUR), 2018, 50(6): 86.
- [5] Woolf V. Character in fiction[J]. The Essays of Virginia Woolf, 1924, 3: 1919-1924.
- [6] Lodge D. The art of fiction[M]. Random House, 2012.
- [7] Forster E M. Aspects of the Novel[M]. Houghton Mifflin Harcourt, 1985.
- [8] Zhang W, Cheung J C K, Oren J. Generating Character Descriptions for Automatic Summarization of Fiction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 7476-7483.
- [9] Bamman D, O'Connor B, Smith N A. Learning latent personas of film characters[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2013, 1: 352-361.
- [10] Bamman D, Underwood T, Smith N A. A bayesian mixed effects model of literary character[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014, 1: 370-379.
- [11] Flekova L, Gurevych I. Personality profiling of fictional characters using sense-level links between lexical resources[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 1805-1816.
- [12] Massey P, Xia P, Bamman D, et al. Annotating character relationships in literary texts[J]. arXiv preprint arXiv:1512.00728, 2015.
- [13] Srivastava S, Chaturvedi S, Mitchell T. Inferring interpersonal relations in narrative summaries[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [14] Iyyer M, Guha A, Chaturvedi S, et al. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 1534-1544.
- [15] Chaturvedi S, Iyyer M, Daume III H. Unsupervised learning of evolving relationships between literary characters[C]//Thirty-First AAAI Conference on Artificial Intelligence. 2017.
- [16] Frermann L, Szarvas G. Inducing semantic micro-clusters from deep multi-view representations of novels[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 1873-1883.
- [17] 赵京胜, 张丽, 朱巧明, 等. 中文文学作品中的社会网络抽取与分析[J]. 中文信息学报, 2017, 31(2): 99-106.
- [18] Chen B, Wang Y. Character interaction network analysis of chinese literary work: a preliminary

- study[C]//Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology. American Society for Information Science, 2016: 88.
- [19] Elson D K, Dames N, McKeown K R. Extracting social networks from literary fiction[C]//Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics, 2010: 138-147.
- [20] Qiu L, Zhang Y. Word segmentation for Chinese novels[C]//Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.
- [21] Bamman D, Popat S, Shen S. An annotated dataset of literary entities[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 2138-2144.
- [22] Muzny G, Fang M, Chang A, et al. A two-stage sieve approach for quote attribution[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. 2017: 460-470.
- [23] De Marneffe M C, Manning C D. Stanford typed dependencies manual[R]. Technical report, Stanford University, 2008.
- [24] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.
- [25] Levy O, Goldberg Y. Dependency-based word embeddings[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014, 2: 302-308.
- [26] Jia Y, Yu S, Chen Z. Chinese word sense induction with basic clustering algorithms[C]//CIPS-SIGHAN Joint Conference on Chinese Language Processing. 2010.
- [27] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

作者联系方式：贾玉祥，河南省郑州市科学大道 100 号郑州大学信息工程学院，450001，15890056535，ieyxjia@zzu.edu.cn