



基于人工智能的文学人物人格分析



朱廷劭
中国科学院心理研究所计算网络心理实验室
Computational CyberPsychology Lab (CCPL),
Institute of Psychology, Chinese Academy of Sciences



艺术来源于生活又高于生活



- 以艺术形象的存在方式区分为：时间艺术（音乐、**文学**）、空间艺术（雕塑、绘画）和时空艺术（戏剧、影视）
- 以主体对艺术的感受方式区分：听觉艺术（如音乐）、视觉艺术（如绘画）和视听艺术（如戏剧）

2018/10/11 Institute of Psychology, CAS 2

文学作品中塑造了很多人物



2018/10/11

Institute of Psychology, CAS

3

分析文学作品中的个性

由于文学人物是**虚拟**的或**理想化**的，**无法**采用传统心理学中自我报告的方法对其进行**直接测量**，主要依赖质性分析或精神分析的手段来作出主观推断，**难以**开展与**现实人生相联系**的人格及文化分析。

2018/10/11

Institute of Psychology, CAS

4

个体心理测量主要依赖于自我报告



自我报告需要用户主动配合。

2018/10/11

Institute of Psychology, CAS

5

专家他评代替自我报告

自我报告完全依赖于用户的配合，如果用户配合度低或者完全不配合，就很难开展。



自我报告



专家他评

根据对被试的行为观察，依据经验，给出心理测量结果。

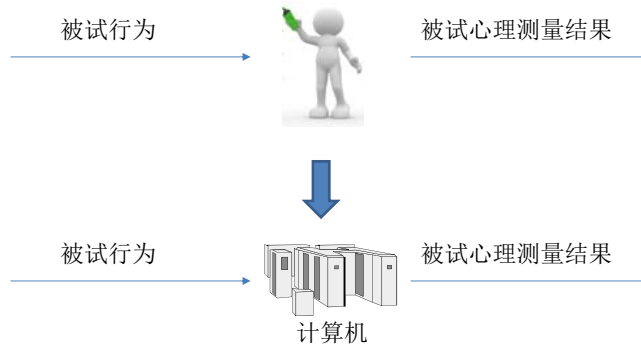
2018/10/11

Institute of Psychology, CAS

6

专家他评的问题

专家他评可以摆脱对被试的依赖，但是实施起来有困难，效率不高，开销大。



2018/10/11

Institute of Psychology, CAS

7

计算机代替专家他评



输入：行为，如何获取行为，如何量化？

输出：心理测量结果

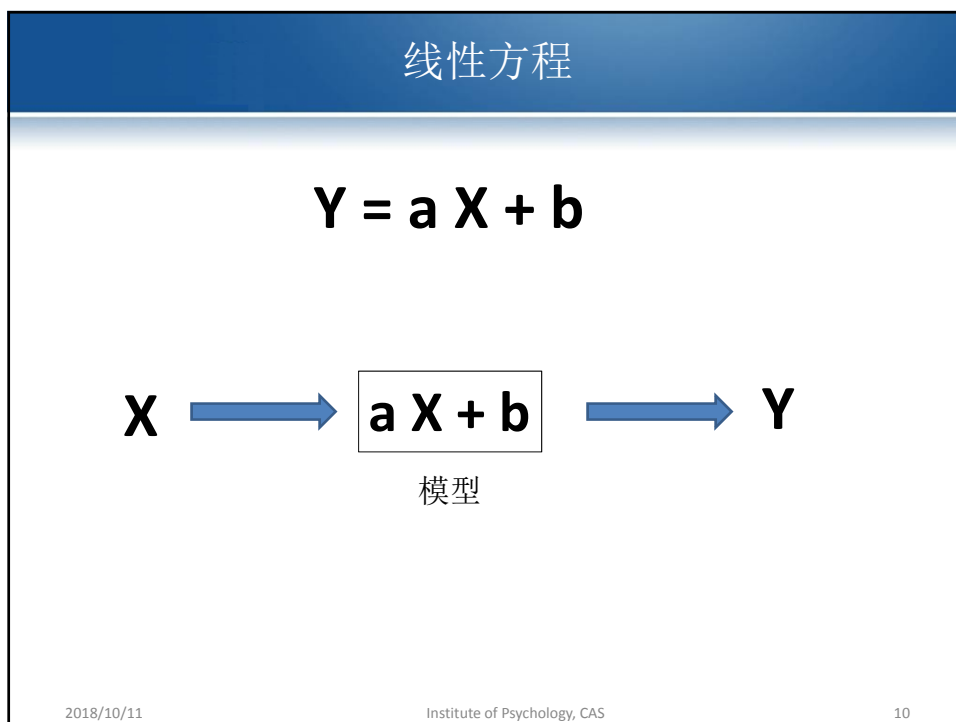
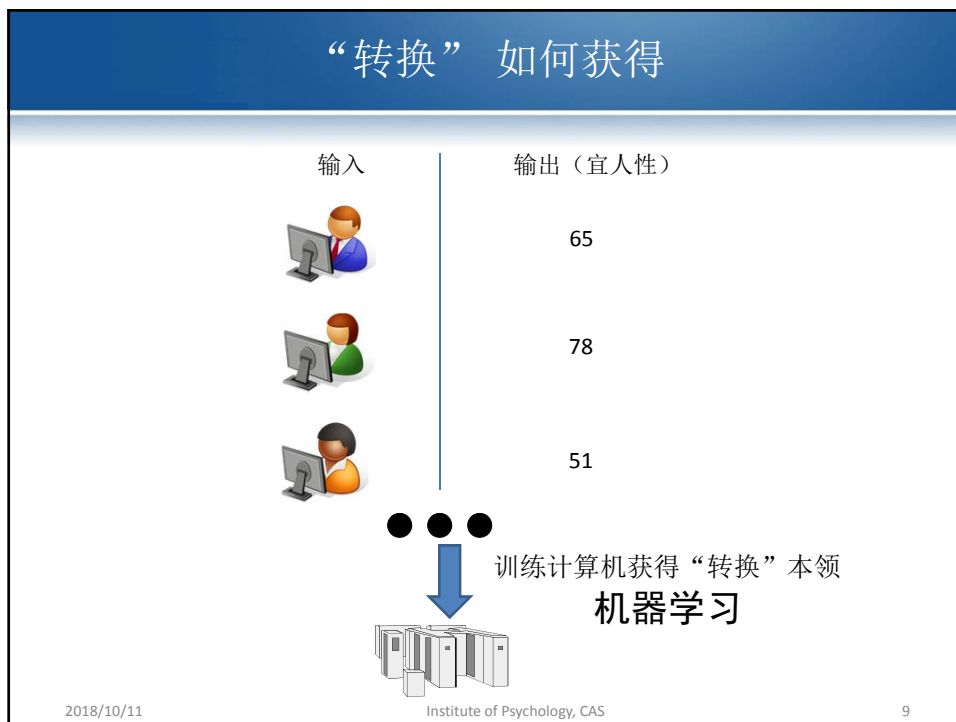
转换：需要将输入转换为输出

输出 = 转换 (输入)

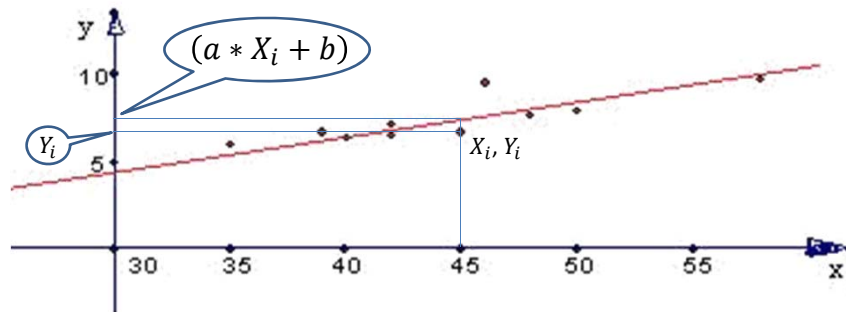
2018/10/11

Institute of Psychology, CAS

8



利用数据训练模型



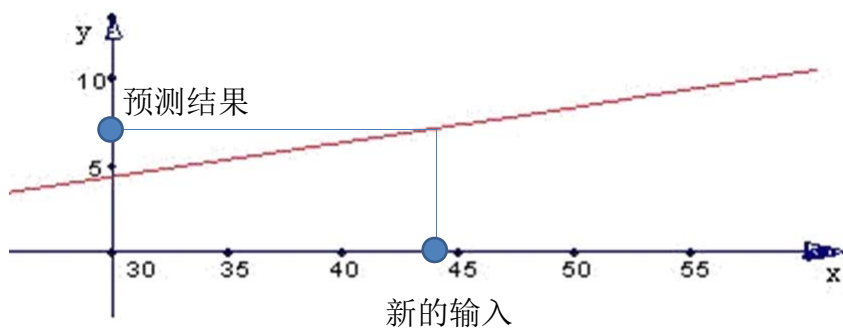
$$\min \sum (Y_i - (a * X_i + b))$$

2018/10/11

Institute of Psychology, CAS

11

预测模型的应用



2018/10/11

Institute of Psychology, CAS

12

机器学习的一般解释

$$\operatorname{argmin} (Y - W(X))$$

原始输出

转换

原始输入

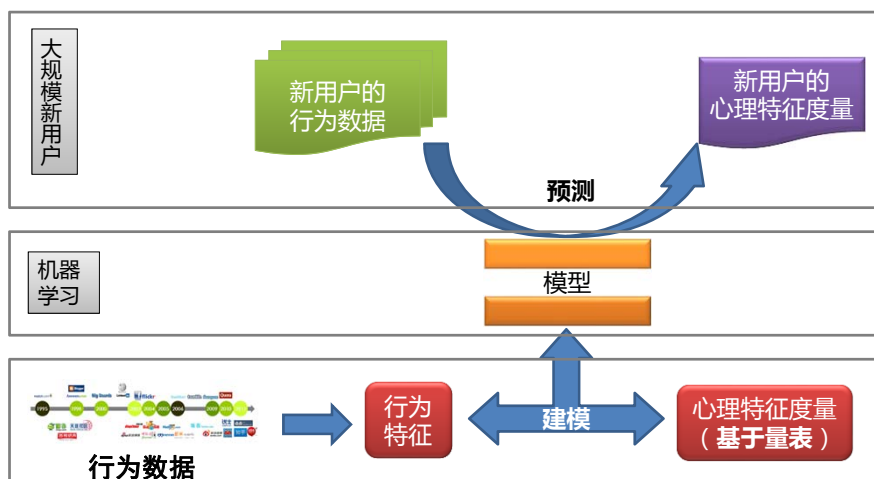
希望得到这样的转换函数（预测模型），它对原始输入转换后的结果，与原始输出之间的误差最小。

2018/10/11

Institute of Psychology, CAS

13

大数据机器学习运用于心理特征预测的两阶段过程



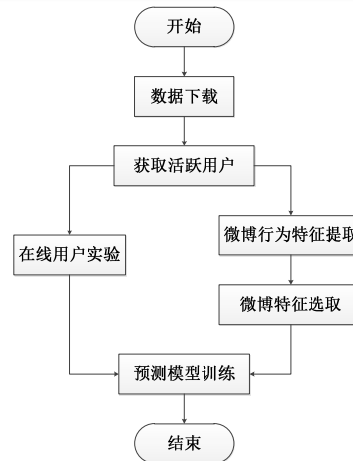
2018/10/11

Institute of Psychology, CAS

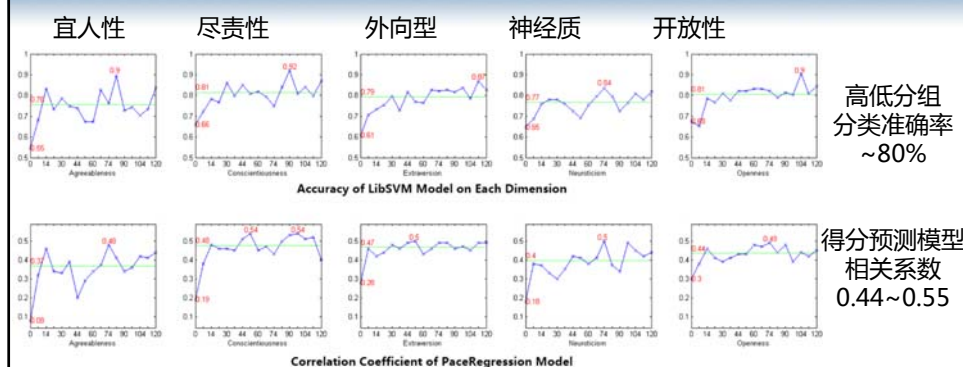
14

如何建立微博用户的人格预测模型

- 从微博平台抓取实验用户数据
- 利用人格量表对用户数据进行标注
- 总结微博特有的行为特征，并以此为依据进行数据的特征提取。
- 利用标注的数据训练出个体模型



微博用户的人格预测



利用微博数据，能够实现对用户人格特征的预测！

Li L, Li A, Hao B, Guan Z, Zhu T (2014) Predicting Active Users' Personality Based on Micro-Blogging Behaviors. PLoS ONE 9(1): e84997. doi:10.1371/journal.pone.0084997

利用行为数据结合机器学习

通过行为数据，利用机器学习训练得到的心理特征预测模型，即生态化识别（Ecological Recognition, ER），能够帮助我们开展很多以往比较难以开展的研究，大大拓展研究的范畴。

利用人工智能和大数据进行文学智能分析，通过生态化识别实现文学人物心理的现实化映射，为文学人物个性研究开辟新的道路。

Mingming Liu, Jia Xue, Nan Zhao, Xuefei Wang, Dongdong Jiao, Tingshao Zhu. (2018). Using Social Media to Explore the Consequences of Domestic Violence on Mental Health. *Journal of Interpersonal Violence*. Feb. 2018, <https://doi.org/10.1177/0886260518757756>

2018/10/11

Institute of Psychology, CAS

17

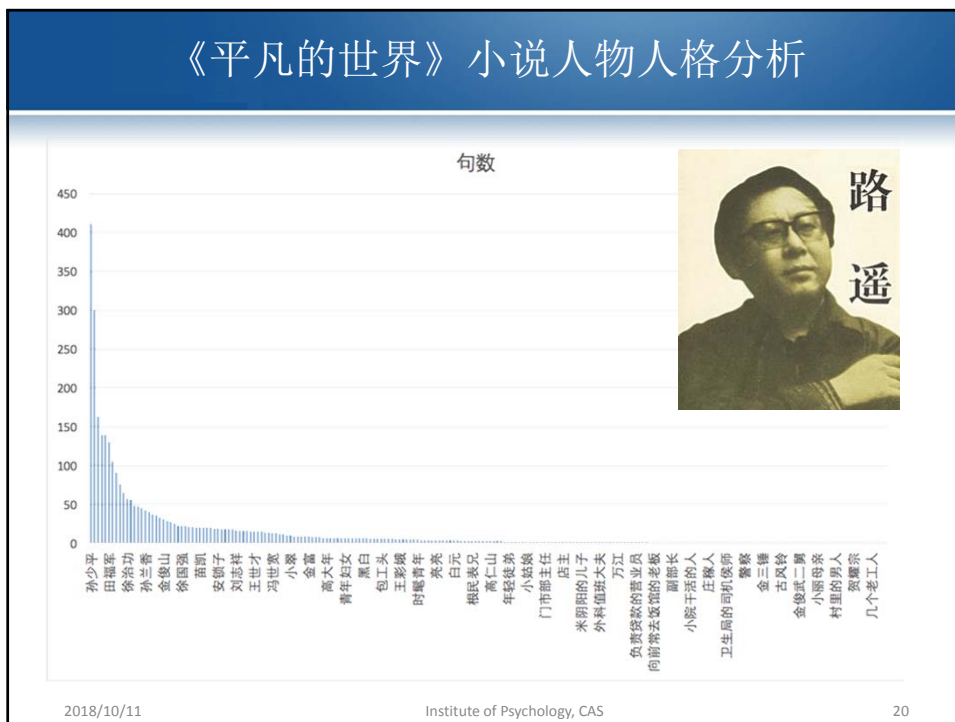
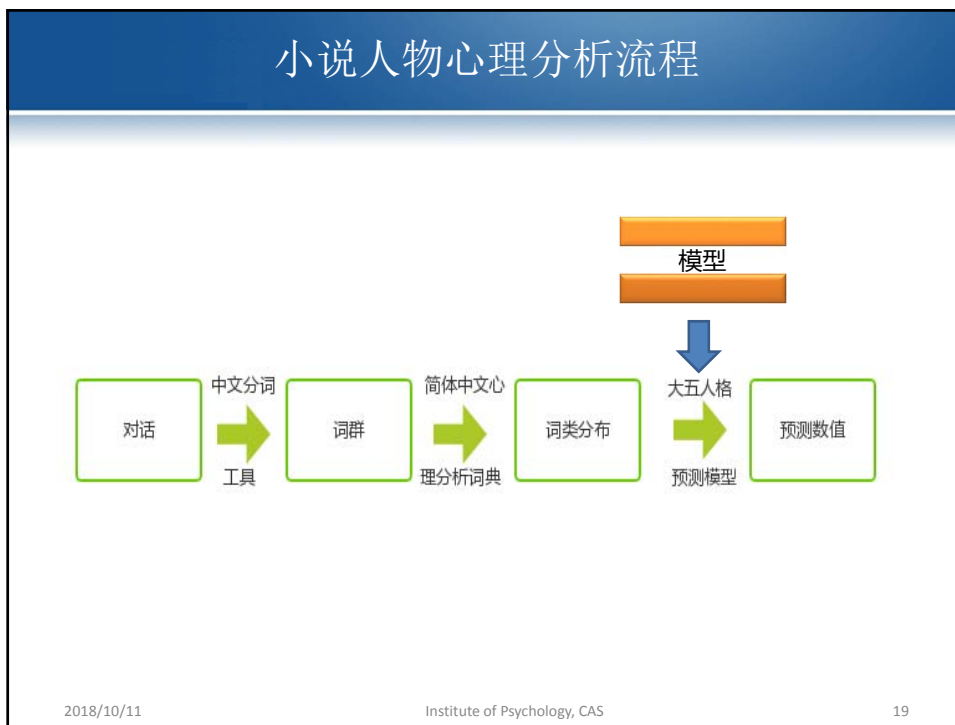
文学智能分析目标



2018/10/11

Institute of Psychology, CAS

18



主要人物的人格预测（改革开放初期）

姓名	A宜人性	C尽责性	E外向性	O开放性	N神经质
孙少安	67.91	84.23	94.13	81.97	75.61
孙少平	59.55	88.81	84.89	96.94	66.79
田润叶	62.66	96.00	95.00	73.76	63.78
田晓霞	72.49	66.08	82.52	94.35	61.03
平均数(标准差)	65.65(4.95)	83.78(11.05)	89.13(5.50)	86.76(9.39)	66.80(5.48)

吴育锋,吴胜涛,朱廷劭,刘洪飞,焦冬冬. 小说人物性格的文学智能分析: 以《平凡的世界》为例[J]. 中文信息学报, 2018, 32(7): 128-136.

2018/10/11

Institute of Psychology, CAS

21

主要人物在重大生活事件前后的人格变化

姓名		A宜人性	C尽责性	E外向性	O开放性	N神经质
孙少安	结婚前	81.27233	85.83807	95.54907	88.55042	60.86471
	结婚后	60.68942	83.37657	93.35589	78.37709	83.58521
孙少平	高中毕业前	50.8893	82.69174	53.24782	89.93696	77.8545
	高中毕业后	62.17855	90.67123	94.46561	99.03313	63.43671

Mingming Liu, Yufeng Wu, Dongdong Jiao, Michael Shengtao Wu, Tingshao Zhu; Literary intelligence analysis of novel protagonists' personality traits and development, Digital Scholarship in the Humanities, , fqy020, <https://doi.org/10.1093/lhc/fqy020>

2018/10/11

Institute of Psychology, CAS

22

金庸小说



“飞雪连天射白鹿，笑书神侠倚碧鸳”（越女传）

2018/10/11

Institute of Psychology, CAS

23

金庸小说创作三阶段

创作阶段	思想特点	人物特点	代表作品
第一阶段	群体本位英雄 儒文化的矛盾 与反复。	崇尚儒家“达则兼济天下” 的胸怀的“入世者”。 如：陈家洛、郭靖、乔峰 等。	《书剑恩仇录》（1955） 《碧血剑》（1956） 《射雕英雄传》（1957） 《神雕侠侣》（1959） 《雪山飞狐》（1959） 《飞狐外传》（1960） 《鸳鸯刀》（1961） 《白马啸西风》（1961）。
第二阶段	个体本位英雄 文化虚无主义， 崇尚道家自然。	崇尚道家的恬淡自然、反 派封建礼教、向往与世无 争的生活的“出世者”。 如：令狐冲、杨过、张无 忌。	《倚天屠龙记》（1961） 《连城诀》（1963） 《天龙八部》（1963） 《侠客行》（1965） 《笑傲江湖》（1969） 《越女剑》（1970）。
第三阶段	反英雄主义 世俗哲学。	追求个性自由、反叛传统 侠义的“玩世者”。 如：韦小宝	《鹿鼎记》（1972）

2018/10/11

Institute of Psychology, CAS

24

各部小说人物对话数量

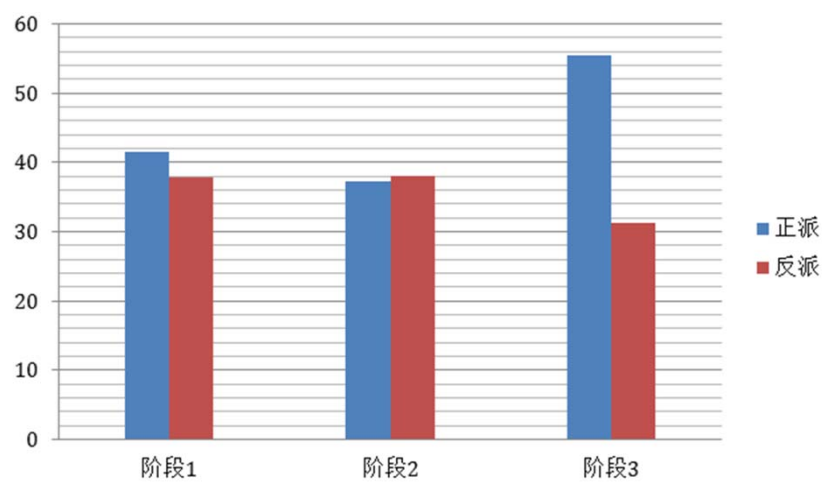
小说	人物数量	人物对话数量
白马啸西风	10	322
碧血剑	39	693
飞狐外传	31	709
连城诀	21	1138
鹿鼎记	113	5060
射雕英雄传	57	1546
神雕侠侣	63	3802
书剑恩仇录	101	2338
天龙八部	214	5346
侠客行	144	1779
笑傲江湖	71	2132
雪山飞狐	16	278
倚天屠龙记	86	2633
鸳鸯刀	11	220

2018/10/11

Institute of Psychology, CAS

25

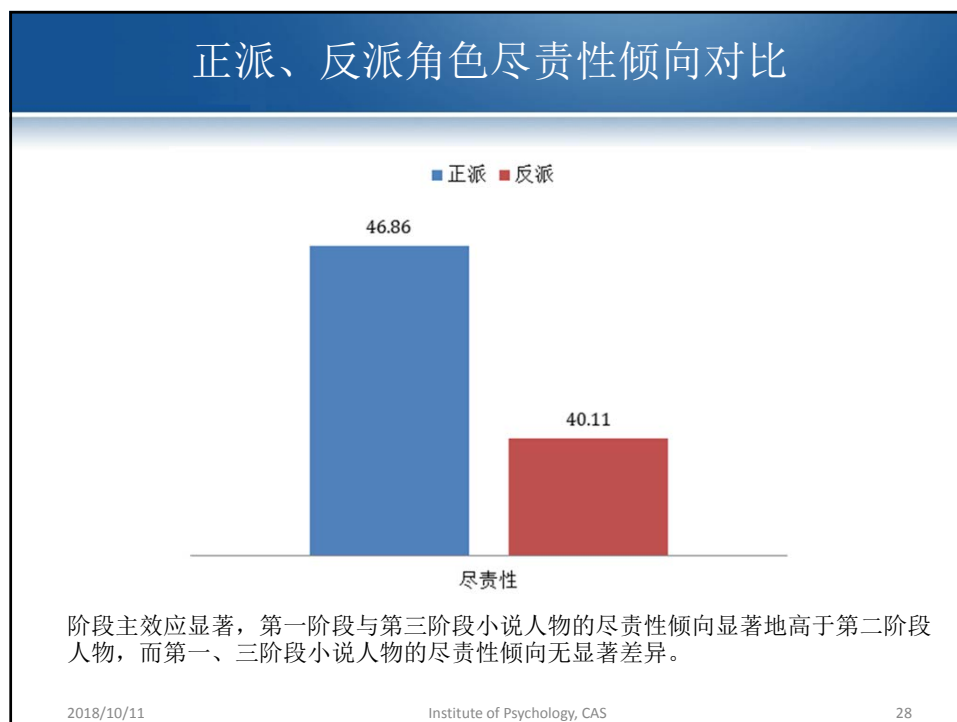
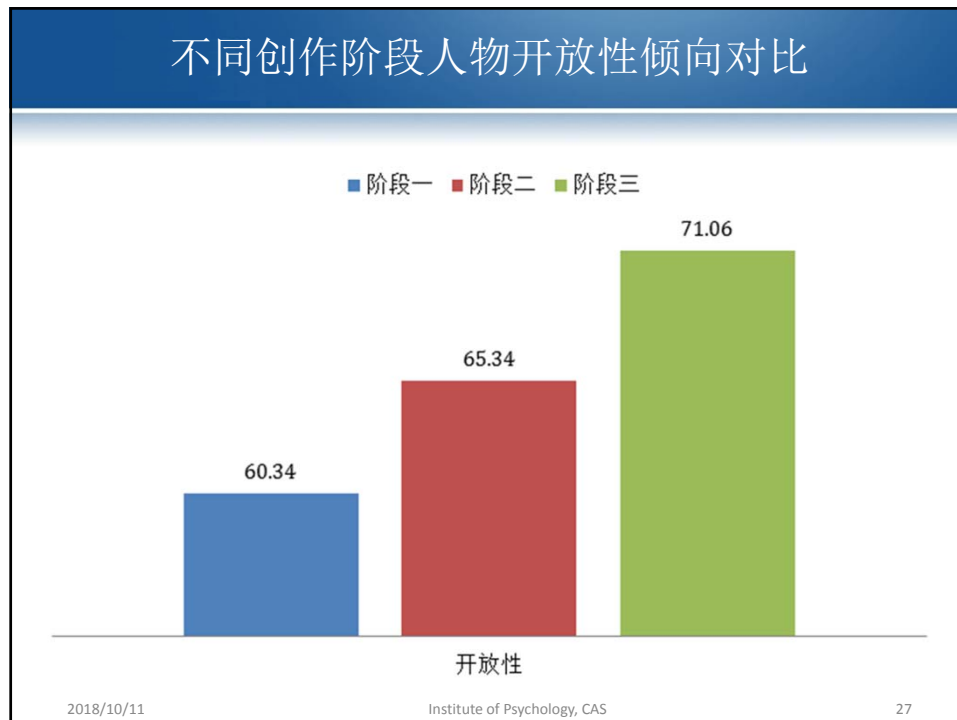
不同创作阶段正派、反派角色宜人性倾向对比



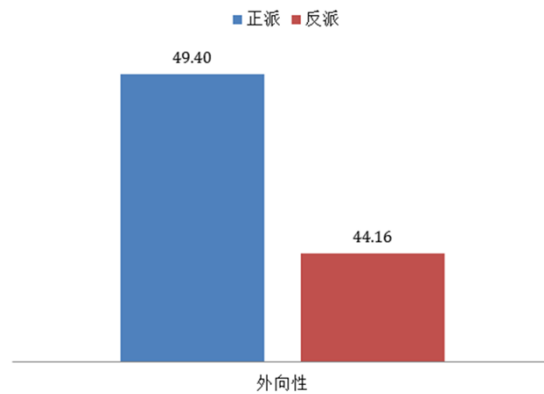
2018/10/11

Institute of Psychology, CAS

26



不同正派、反派角色外向性倾向对比



阶段主效应不显著

不同正派、反派角色神经质对比

在神经质倾向上，派别主效应（ $M_{\text{正}}=59.09$ vs $M_{\text{反}}=60.07$, $F(1, 737)=0.186$, $p=0.667>0.05$)、阶段主效应（ $M_1=60.47$ vs $M_2=59.42$ vs $M_3=58.85$, $F(2, 736)=0.18$, $p=0.83>0.05$)和派别与阶段的交互作用（ $F(2, 736)=1.61$, $p=0.20>0.05$)均不显著。

金庸创作过程分析

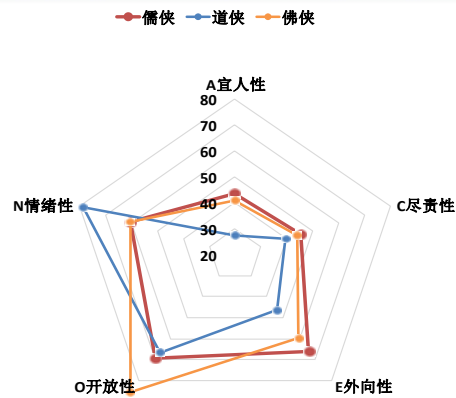
- 在金庸小说创作的第二、三阶段，小说人物的开放性倾向显著地高于第一阶段人物，体现了金庸创作早期塑造的人物更加倾向于遵守传统、常规礼教、不善变通的性格特点，而其后期塑造的人物则更具独创性，具有想象活跃、更易接受新思想的特点。
- 随着年龄的增长与人生阅历的增加，金庸从崇尚传统“儒家”思想到开始追求“道家”、“佛家”精神，对于封建礼教与常规化的、定势的社会成规表现出了更强的反叛精神，而对“侠义”的解读更是体现出了极强的包容性，其笔下开始出现许多“新武侠”或“反武侠”人物。

崔馨月,郑苑仪,王贝依,郭荣慧,丁照云,朱廷劭.金庸小说人物的大五人格—基于文心系统的15部金庸小说分析.[ChinaXiv:201802.00674]

武侠的儒、释、道分类

人物	小说	句数	身份信息	
儒侠	陈家洛	书剑恩仇录	505	红花会总舵主
	徐天宏	书剑恩仇录	158	红花会七当家
	袁承志	碧血剑	264	武林盟主
	郭靖	射雕英雄传	247	北侠、武林盟主
	陈近南	鹿鼎记	104	天地会总舵主
道侠	周伯通	射雕英雄传	65	全真派“老顽童”
	黄药师	射雕英雄传	60	桃花岛岛主
	杨过	神雕侠侣	1056	神雕大侠
	胡斐	飞狐外传	264	雪山飞狐
	令狐冲	笑傲江湖	590	华山派大弟子
佛侠	张翠山	倚天屠龙记	168	武当七侠（排行第五）
	洪七公	射雕英雄传	90	丐帮帮主
	张无忌	倚天屠龙记	632	明教教主
	谢逊	倚天屠龙记	116	明教四大护教法王之一
	狄云	连城诀	312	悲苦侠客
	段誉	天龙八部	770	大理国皇帝
	乔峰	天龙八部	354	丐帮帮主
	虚竹	天龙八部	181	小僧
	石破天	侠客行	439	长乐帮主
方证	笑傲江湖	63	少林派掌门	

不同文化流派武侠小说人物的大五人格



吴胜涛, 蔡佳佳, 黄议莹, 刘明明, 朱廷劭. 基于人工智能的小说人物心理分析. 【在审】

2018/10/11

Institute of Psychology, CAS

33

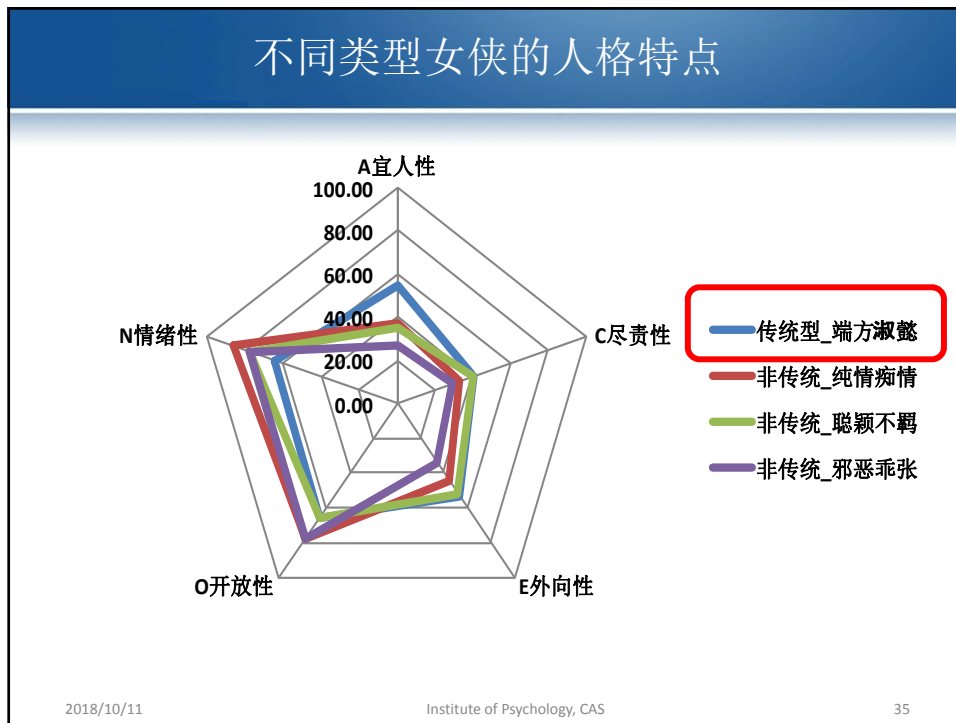
不同文化流派武侠小说人物的大五人格

女侠类别	姓名	句数	年龄	地域	家庭	婚姻
纯情痴情	小龙女	310	17-39	终南山活死人墓	孤儿	已婚
	王语嫣	190	18-21	姑苏曼陀山庄	镇南王私生女	未明
	钟灵	101	16	大理万劫谷	镇南王私生女	已婚
	沐剑屏	136	14	云南沐王府	沐英后代	已婚
	木婉清	167	少女	深山	镇南王私生女	已婚
	阿朱	329	15-17	姑苏听香水榭	镇南王私生女	已婚
聪颖不羁	赵敏	179	17	蒙古	汝阳王之女	已婚
	郭芙	163	9-31	襄阳、桃花岛	大侠之后	已婚
	郭襄	166	15	襄阳	大侠之后	未婚
	陆无双	102	9	江南嘉兴	庄主千金, 灭门之灾	未婚
	霍青桐	140	19	天山	首领之女	未明
邪恶乖张	李莫愁	168	30-37	终南山、赤霞山庄	未明	未婚
	阿紫	316	15	小镜湖	镇南王私生女	未婚
	马夫人	127	27	中原山村	遗孀	已婚
端方淑懿	双儿	119	14	湖州庄家府邸	孤儿	已婚
	戚芳	101	18	湘西乡下	生父	已婚

2018/10/11

Institute of Psychology, CAS

34

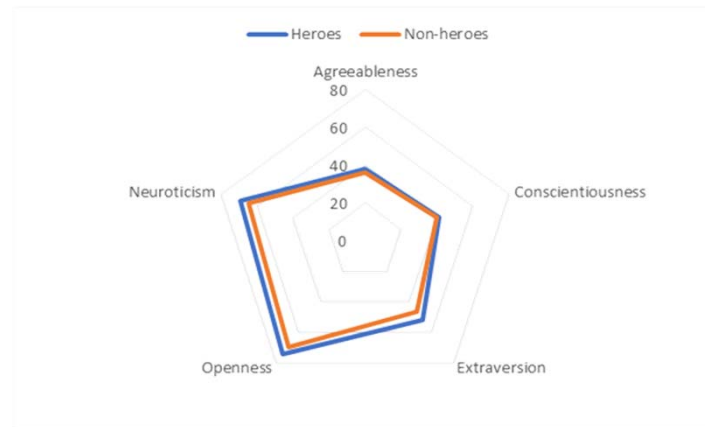


英雄理想人物

陈家洛	清朝大学士陈仕官之子，汉族，胞弟乾隆（爱新觉罗·弘历），红花会总舵主，秀才，约20岁，游侠
袁承志	山海关总兵袁崇焕之子，汉族，七省武林盟主，闯王李自成的军事与政治盟友，流浪孤儿-被师傅穆人清收养长大，游侠
郭靖	梁山好汉“地佑星”郭盛后人，汉族，武林领袖（郭大侠），襄阳城防务领袖，大侠
杨过	杨家将后人，汉族，父亲杨康因背叛结义兄长勾结金人为人不齿，游侠，流浪孤儿
狄云	出身不详，游侠，没有突出的政治贡献，游侠
张无忌	武当侠客张翠山后人，汉族，中土明教34代教主，武当张三丰传人，反元势力领袖，归隐侠士
段誉	大理太子段延庆之子，镇南王段正淳养子，汉族，大理皇位第一继承人，大理宣仁帝，退位后避位为僧，涉足江湖时期为游侠
虚竹	少林伏虎罗汉玄慈方丈之子，汉族，逍遥派传人，西夏国驸马，游侠
萧峰	辽国珊军总教头萧远山之子，契丹族，丐帮帮主，辽国南院大王（楚王），游侠，军事领袖，政客
石破天	玄素住主石清之子，梅芳姑养子，长乐帮主，武林盟主，游侠
令狐冲	岳不群养子，华山派大弟子，恒山派掌门，归隐游侠，音乐家
胡一刀	李自成四大护卫之飞天狐狸后人，辽东大侠
苗人凤	李自成四大护卫之一的后人，游侠
张翠山	张三丰养子，武当七侠排行第五，游侠
胡斐	李自成四大护卫之飞天狐狸后人，大侠胡一刀之子，游侠
洪七公	丐帮帮主，四大高手之一，游侠
文泰来	红花会四当家，游侠
曲阳	日月神教长老，音乐家，游侠
刘正风	衡山派长老，后拜官参将，音乐家，游侠
莫大	衡山派掌门，音乐家，游侠，隐士
陈近南	天地会创办人、总舵主，韦小宝师傅，政治、军事领袖（反清复明运动），游侠

2018/10/11 Institute of Psychology, CAS 36

对比结果



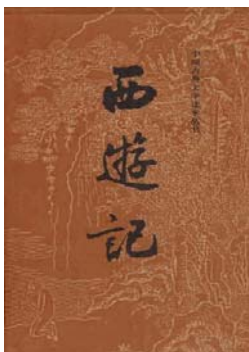
Yijun Yang, Michael Shengtao Wu, Nan Zhao, Tingshao Zhu. Identifying Personality Traits of Chinese Swordsmen 【在审】

2018/10/11

Institute of Psychology, CAS

37

“西游记”大家都读过

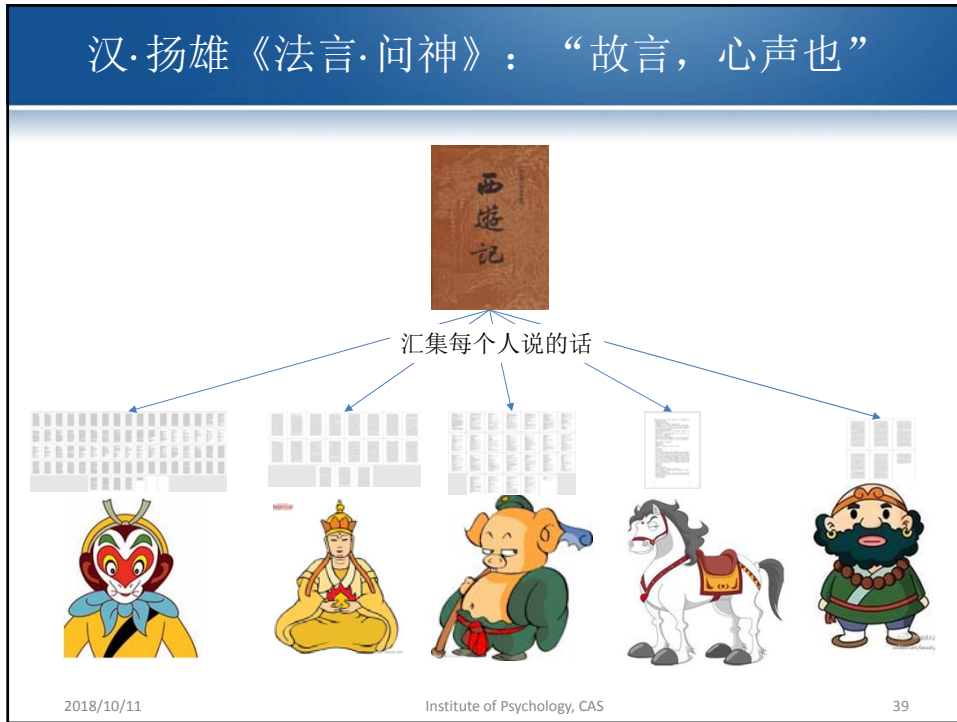


2018/10/11

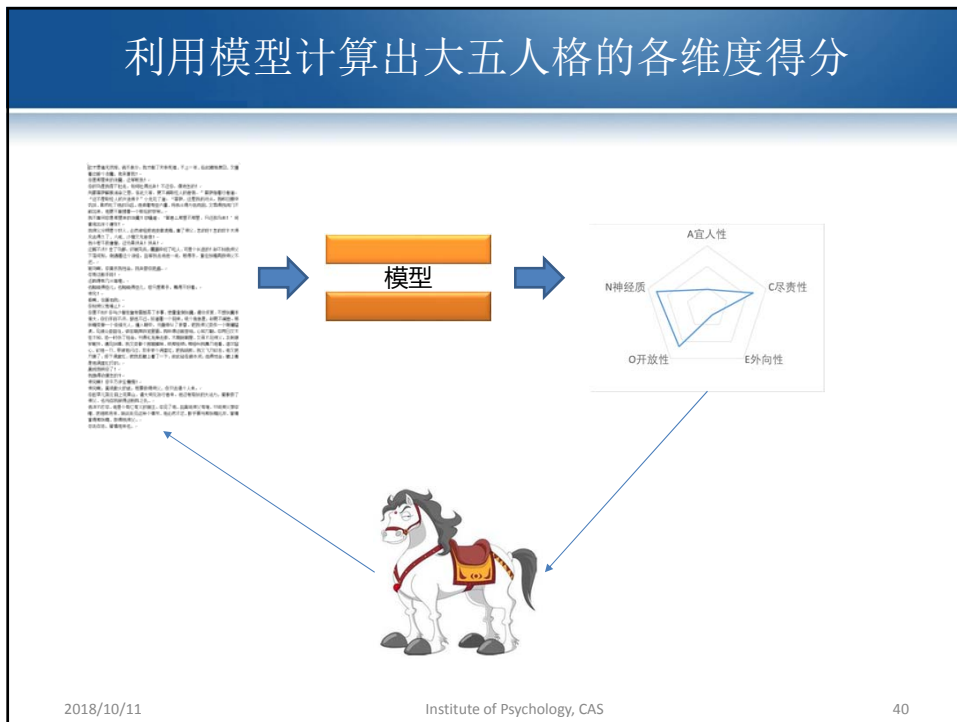
Institute of Psychology, CAS

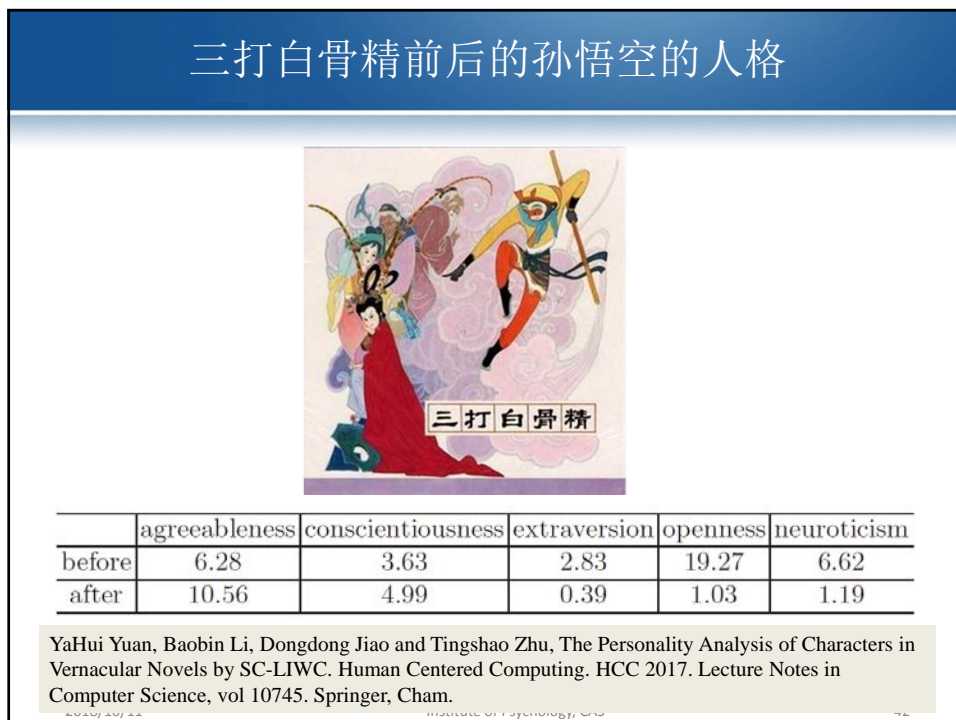
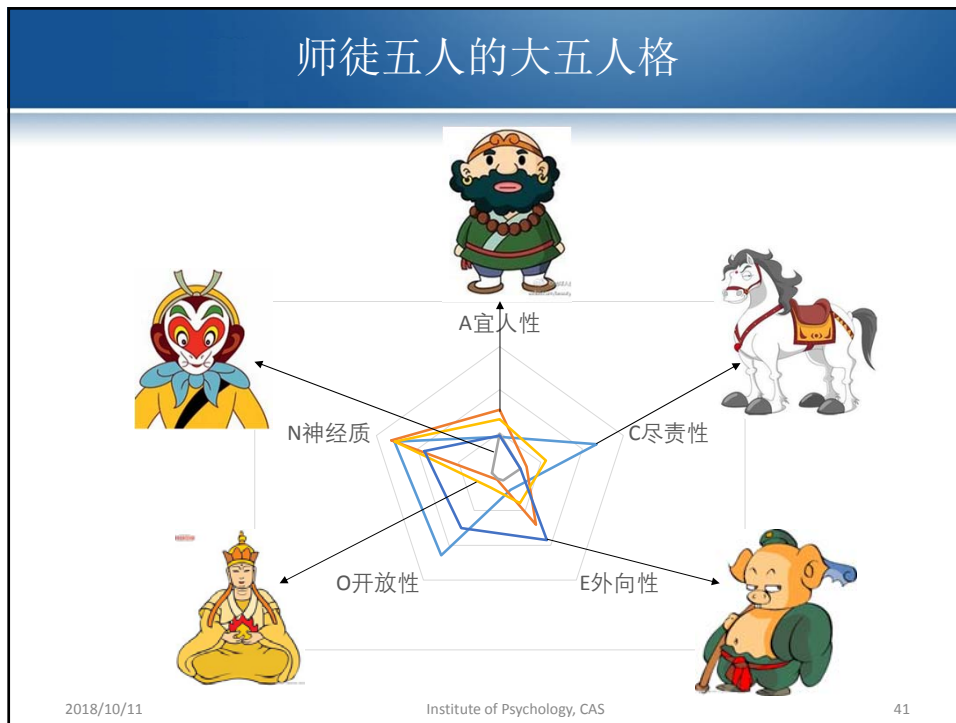
38

汉·扬雄《法言·问神》：“故言，心声也”



利用模型计算出大五人格的各维度得分





总结

- 大数据为心理学的研究和应用带来一个新的机遇；
- 文学智能分析实现对人物的个性研究；
- 对传统文化中的个性进行心理学解析；
- 不同时代背景下人物个性特点的分析；
- 利用文言文进行历史人物心理特点以及纵向分析！

2018/10/11

Institute of Psychology, CAS

43

谢谢！

Thank You!



中国科学院心理研究所计算网络心理实验室
Computational CyberPsychology Lab. (CCPL),
Institute of Psychology, Chinese Academy of Sciences

<http://ccpl.psych.ac.cn>