

# Report of Deep Learning for Natural Language Processing 3

## Use Word2Vec model to train word vector

Jiayi Zhang  
zy2303814@buaa.edu.cn

### Abstract

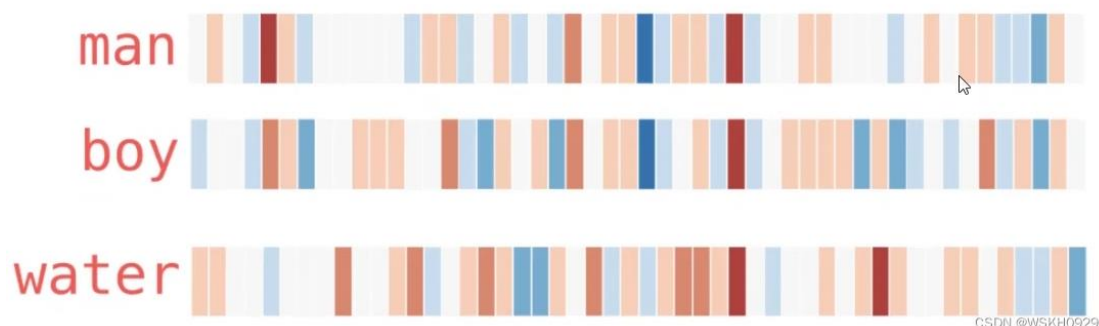
In this assignment, the Word2vec model is trained by using Jin Yong's 16 novel texts to generate word vectors of processed words, and the effectiveness of the trained word vectors is verified by calculating the semantic distance between the word vectors, the clustering results of the same type of words, and the text prediction.

### Introduction

There are two common neural networks used to train language models: CBOW(continuous bag of words) predicts the current word with context and skip-gram predicts the context with current word. It is generally composed of input layer, hidden layer and output layer.

Word2vec is a popular technique used in natural language processing (NLP) to represent words in a continuous vector space. Developed by a team of researchers led by Tomas Mikolov at Google, Word2vec has become a fundamental tool in many NLP applications.

Word2vec is a group of related models that are used to produce word embeddings, which are dense vector representations of words. These embeddings capture semantic relationships between words, such that words with similar meanings are represented by similar vectors. In general, the higher the dimension of a vector, the more information it can provide, and the more reliable the results are. As can be seen from **Figure 1**, the generated word vectors are represented by heat maps, and similar words are more similar in the feature expression, thus proving that the construction of word vectors is meaningful.



**Figure 1** Representation of a vector of similar words[1]

### Methodology

#### M1: Data Preprocessing

First of all, it is necessary to establish a stop word list. Since people's understanding of novels

starts with nouns, such as personal names and place names, it is necessary to remove many meaningless words. The stop word list takes into account special characters, meaningless personal pronouns such as “他”、“她”, as well as words such as “一下”“一个”“一些”and so on.

## M2: Training the Word2vec model

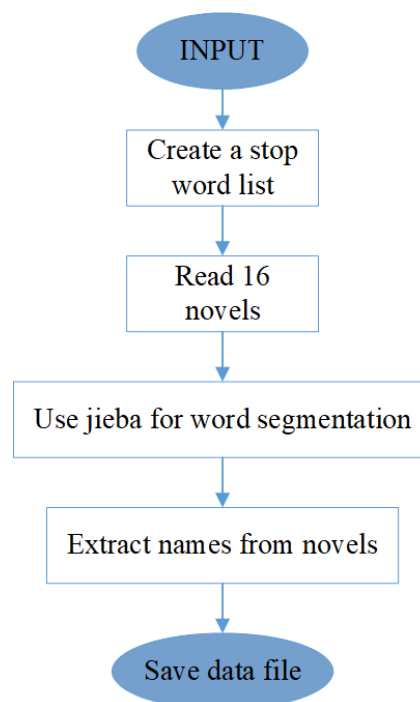
Word2Vec is a popular word embedding technique that converts words into a vector form to capture semantic relationships between words. Text generation with Word2Vec usually involves the following steps:

- Preprocessed text
- Training the Word2Vec model
- Choose a generation strategy: Determine how to use word vectors to generate text. Common strategies include:
  - 1) Continuity: The word vector of the previous words in the text is used to predict the next word.
  - 2) Average strategy: average all word vectors in a piece of text, and then use this average vector to predict the next word.
  - 3) Contextual relevance: Choose the word that is most relevant to the context.
- Generated text
- 4) Initialize context: Select one or more words as the starting point for generating text.
- 5) Predict the next word: Use the Word2Vec model and selected strategies to predict the next word.
- 6) Sentence construction: Repeat the prediction process until a complete sentence is generated or a predetermined number of words is reached.

## Experimental Studies

### Step1: Data Preprocessing

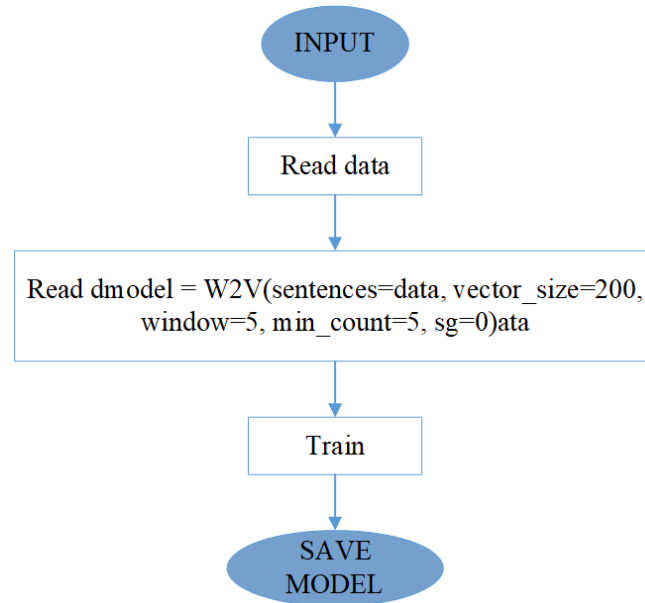
The first step is data preprocessing, and the algorithm flow chart is shown in **Figure 2**.



**Figure 2** Data preprocessing algorithm

### Step2: Word2vec model training

The second part conducts model training, and the algorithm flow chart is shown in **Figure 3**. After the model is trained, the model verification process can be entered.



**Figure 3** Training model algorithm

### Step3: Text similarity verification

Firstly, the model is used to calculate the similarity of the text, and then the validity of the constructed word vector is verified. There are many classic names in Mr. Jin Yong's novels, between them or do not know each other, or have intimate relationships, such as "张无忌" and "赵敏", "周芷若" have a love relationship, "乔峰" and "段誉" are friends, "郭靖" and "黄蓉", "杨过" and "小龙女" are husband and wife relationship. Therefore, this experiment first verifies the Euclidean distance and cosine similarity between these names, and the results are shown in Table1 and Table2.

**Table 1** Euclidean distance between word vectors

Euclidean distance	张无忌	乔峰	郭靖	杨过	令狐冲	韦小宝
张无忌	0	8.134056	9.777371	9.229698	7.369474	11.94606
赵敏	6.86718	5.898668	8.14549	8.129938	7.360621	11.20477
周芷若	6.82691	5.538126	7.77546	7.830071	7.568213	11.6982
段誉	8.371193	6.270872	8.58105	7.578622	7.736325	12.05796
黄蓉	9.399901	8.291562	5.102507	5.770179	9.748766	11.31764
小龙女	9.283319	7.565599	7.288921	5.136044	9.39287	12.90974
岳不群	8.106502	5.269957	7.918685	8.343549	6.642999	12.63532
东方不败	9.07307	5.120127	9.286642	9.593461	8.983706	12.26362
鳌拜	11.02441	6.94202	10.79947	11.22215	11.2946	12.46185

Table 2 Cosine similarity between word vectors

Cosine similarity	张无忌	乔峰	郭靖	杨过	令狐冲	韦小宝
张无忌	1	0.654346	0.569523	0.62949	0.765886	0.511387
赵敏	0.776733	0.645801	0.618052	0.649957	0.734052	0.516143
周芷若	0.784698	0.675069	0.656637	0.681683	0.716215	0.453064
段誉	0.671338	0.771611	0.636228	0.729422	0.721255	0.473774
黄蓉	0.601418	0.595038	0.877185	0.849363	0.572842	0.552448
小龙女	0.6063	0.665445	0.745839	0.879724	0.59851	0.403511
岳不群	0.65898	0.739648	0.647125	0.631064	0.792284	0.348698
东方不败	0.56287	0.656126	0.434425	0.444341	0.587414	0.346817
鳌拜	0.258584	0.452293	0.236371	0.21914	0.220258	0.339643

Table 1 calculates the Euclidean distance between names. **The smaller the distance, the greater the similarity between words.** As can be seen from the numbers marked in red, the word vectors can indeed reflect the familiar relationships between people, but as can be seen from the numbers in blue, calculating the Euclidean distance between the word vectors is sometimes not very accurate.

Table 2 calculates the cosine similarity between people's names. **The greater the value, the greater the similarity between words.** It can be found that calculating the cosine similarity of word vectors can more accurately reflect the relationship between words than calculating the Euclidean distance. Indeed, the word vectors of people with similar relationships that we know subjectively are more similar.

#### Step 4 : Text prediction validation

Secondly, text prediction is carried out by using the established model. After extracting a paragraph from the pre-processed text file, calculating the average value of the context vector, traversing all the words in the lexicon, calculating their similarity with the average vector, filtering out the words that are irrelevant to the context semantics, and selecting the next word using Top-k and temperature sampling. It is also important to consider that the word should not be repeated.

The experiment **predicts 2, 5, 10, and 15 words backwards** by giving text of different lengths (the model training time for more words is too long). The text **in bold** and **blue is the words in the original text**, and **the text in red is the predicted words**. Although different from the original text, the prediction results are still logical, so the validity of the word vector can be proved.

#### Predicted results:

Original text1 = ['男孩', '不去理', '脸上', '正气凛然', '指着', '苗夫人', '没良心', '田归农', '提起', '长剑', '正要', '分心', '刺去', '苗夫人', '一声', '掩面', '哭', '大雨', '中', '直奔', '田归农', '顾不得', '杀', '男孩', '提剑', '追出', '一窜', '一跃', '追到', '苗夫人', '身旁', '劝道', '兰妹', '叫化', '胡说八道', '别理', '苗夫人', '哽咽', '道确', '良心', '不好', '哭', '说话', '脚下', '丝毫', '不停', '田归农', '伸手', '挽', '臂膀', '苗夫人', '用力']

Predict 2 words = [男孩 不去理 脸上 正气凛然 指着 苗夫人 没良心 田归农 提起 长剑 正要 分心

刺去 苗夫人 一声 掩面 哭 大雨 中 直奔 田归农 顾不得 杀 男孩 提剑 追出 一窜 一跃 追到 苗夫人 身旁 劝道 兰妹 叫化 胡说八道 别理 苗夫人 **胡斐右 肋下**]

Predict 5 words = [男孩 不去理 脸上 正气凛然 指着 苗夫人 没良心 田归农 提起 长剑 正要 分心 刺去 苗夫人 一声 掩面 哭 大雨 中 直奔 田归农 顾不得 杀 男孩 提剑 追出 一窜 一跃 追到 苗夫人 身旁 劝道 兰妹 叫化 胡说八道 别理 苗夫人 **执剑 一拉 右腕 手肘 翻过**]

Predict 10 words = [男孩 不去理 脸上 正气凛然 指着 苗夫人 没良心 田归农 提起 长剑 正要 分心 刺去 苗夫人 一声 掩面 哭 大雨 中 直奔 田归农 顾不得 杀 男孩 提剑 追出 一窜 一跃 追到 苗夫人 身旁 劝道 兰妹 叫化 胡说八道 别理 苗夫人 **领口 左腕 揽住 格开 一拉 架开 心口 抱住 左拳 甩脱**]

---

Original text2 = ['男孩', '没想到', '居然', '会肯', '接过', '玉镯', '再要', '回', '李文秀', '男孩', '於', '天铃鸟', '递', '李文秀', '双手', '合著', '鸟儿', '手掌', '中', '感觉', '柔软']

Predict 2 words = [男孩 没想到 居然 会肯 接过 玉镯 **递给 挽住**]

Predict 5 words = [男孩 没想到 居然 会肯 接过 玉镯 **提起 格开 一扯 挽住 托住**]

Predict 10 words = [男孩 没想到 居然 会肯 接过 玉镯 **持着 格开 揪住 后领 胸部 一扯 捏住 托住 一拉 甩出**]

---

Original text3 = ['令狐冲', '倒退', '群豪', '焦急', '耳', '听', '桃谷六仙', '一旁', '胡言乱语', '更增', '恼怒', '令狐冲', '退一步', '波', '一声', '左足', '踏入', '水坑', '心念', '一动', '风', '太师', '叔', '当日', '谆谆教导', '天下', '武术', '千变万化', '神而明之', '存乎一心', '招式', '精妙', '招', '破绽', '独孤', '大侠', '传下来', '路剑法', '打遍天下', '无敌手', '敌招', '之中', '瞧', '破绽', '眼前', '位', '前辈', '剑法', '圆转', '如意', '半分', '破绽', '瞧', '不出', '破绽', '未必', '真', '破绽', '瞧', '不出', '退', '几步', '凝视', '剑光', '所幻', '无数', '圆圈', '蓦地', '心想', '说不定', '圆圈', '中心', '便是', '破绽', '破绽', '一剑', '刺入', '长剑', '绞', '手臂', '登时', '断', '想', '幸好', '攻逼', '只能', '渐进', '要伤', '性命', '不易', '一味', '退避', '终究', '输', '此仗', '一败', '大伙儿', '心虚', '气馁', '闯', '少林', '救', '盈盈', '想到', '盈盈', '情深义重', '断送', '一条', '手臂', '内心深处', '断送', '一条', '手臂', '乃是', '快慰', '之事', '觉', '负', '良多', '须', '伤残', '稍报', '深恩', '言念', '及此', '内心深处', '倒似', '渴望', '一条', '手臂', '斩断', '手臂', '一伸', '长剑', '老者', '剑', '光圈', '中刺', '一声', '响', '令狐冲', '只感', '胸口', '剧烈', '一震', '气血', '翻涌', '一只', '手臂', '完好', '老者', '退', '开', '两步', '收剑', '而立', '脸上', '神色', '古怪', '既有', '惊诧', '之意', '惭愧', '之色', '几分', '惋惜', '之情', '隔', '良久', '令狐公子', '剑法', '高明', '胆识过人', '佩服', '佩服', '令狐冲', '方知', '适才', '冒险', '一击', '找到', '剑法', '弱点', '老者', '剑法', '实在', '太高', '光圈', '中心', '本是', '凶险', '处', '练得', '破绽', '藏于', '天下', '成千成万', '剑客', '之中', '胆敢', '以身', '犯险', '一逞', '暗叫', '侥幸', '侥幸', '一道道', '汗水', '背脊', '流下', '躬身', '前辈', '剑法', '通神', '承蒙指教', '晚辈', '得益', '非浅', '句', '话', '倒', '寻常', '客套', '一战', '武功', '进益', '确是', '大有', '好处', '令', '得知', '敌人', '招数', '中', '最强', '处', '便是', '最弱', '处', '最强', '处', '击破', '迎刃而解', '高手', '比剑', '一招', '而决', '老者', '即见', '令狐冲', '剑', '光圈', '中挥刀', '直入', '令狐冲', '凝视', '半晌', '令狐公子', '老朽', '几句话', '令狐冲', '恭', '聆', '前辈', '教诲', '老者', '长剑', '交给', '挑菜', '汉子', '往东', '走', '令狐冲', '长剑', '抛', '地下', '跟随', '一棵', '大树', '旁', '和群豪', '相去', '数十丈', '虽可', '望见', '话声', '老者', '树荫下', '坐', '指着', '树旁', '一块', '圆石', '请坐下', '说话', '令狐冲', '坐', '缓缓', '令狐公子', '年轻', '一辈', '人物', '之中', '如般', '人才', '武功', '少有', '令狐冲', '晚辈', '行为不端', '声名狼藉', '不容', '师门', '怎配', '承', '前辈', '见重', '老者', '道辈', '武人']

Predict 5 words = [令狐冲 倒退 群豪 焦急 耳 听 桃谷六仙 一旁 胡言乱语 更增 恼怒 令狐冲 退一步 波 一声 左足 踏入 水坑 心念 一动 风 太师 叔 当日 谆谆教导 天下 武术 千变万化 神而明之 存乎一心 招式 精妙.....(中间省略)... 即见 令狐冲 剑 光圈 中挥刀 直入 令狐冲 凝视 半晌 令狐公子 老朽 几句话 令狐冲 恭 聆 前辈 教诲 老者 长剑 交给 挑菜 汉子 往东 走 令狐冲 长剑 抛 地下 跟随 一棵 大树 旁 和群豪 相去 数十丈 虽可 望见 话声 老者 树荫下 坐 指着 树旁 一块 圆石 请坐下 说话 **盛怒之下 腕底 中伤 指力 银丝软**]

---

### Step 5: Text clustering validation

In order to verify the validity of the word vector more intuitively, in the data processing, we use the list of all characters in Jin Yong's novels "金庸小说全人物.txt" to extract a list of characters in each novel (see Appendix), and use the trained model to cluster names in these character lists. Intuitively observe whether characters belonging to the same novel have closer word vectors, and the clustering results are shown in Figure 4.

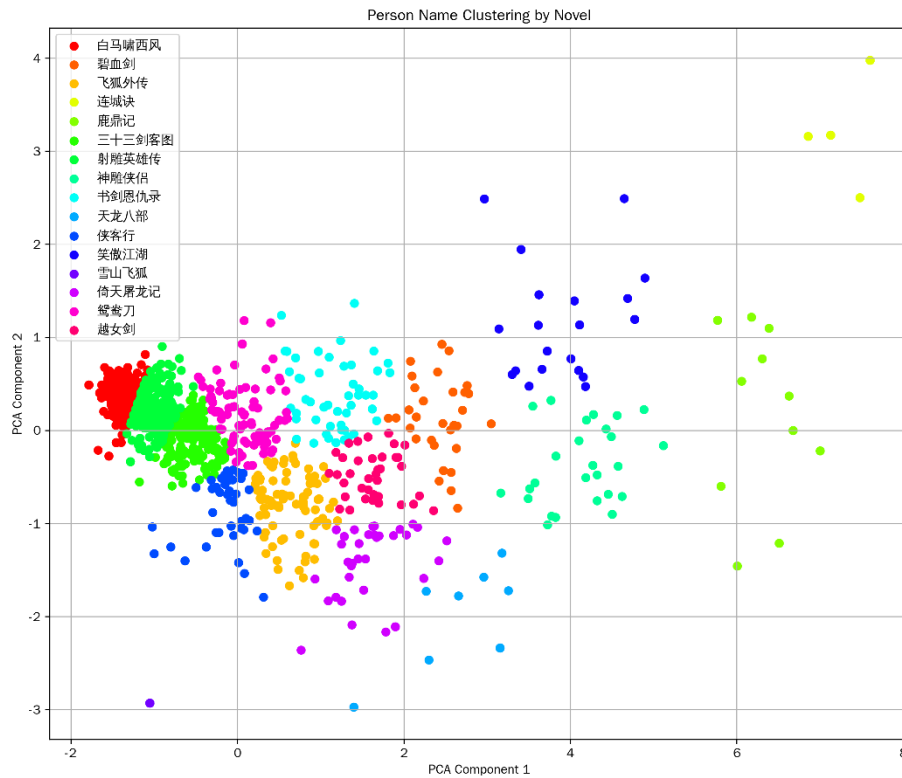


Figure 4 Results of word clustering experiment

It can be seen from the figure that different colors are **successfully clustered**, indicating that the word vectors of characters belonging to the same novel are more similar.

However, it can be noted that the points in the right half of the figure are relatively discrete, indicating that the points near each other do not necessarily belong to the same novel, which can also explain the problem in the test of the word vectors of characters in Step3.

## Conclusions

Word2Vec is a model to learn semantic knowledge in an unsupervised way from a large number of text corpus. It represents the semantic information of words in the form of word vectors, that is, through an embedding space, semantically similar words are close together in the space. It is a mapping of the word from the original space to the new multidimensional space, that is, Embedding the original space of the word into a new space.

In this experiment, a Word2vec model for words in Jin Yong's 16 novels is established. The validity of word vectors is verified by calculating the semantic distance between word vectors, the clustering of personal names, and the use of word vectors for text prediction.

## References

- [1] CSDN. [EB/OL] <http://t.csdnimg.cn/XiNOF>  
[2] CSDN. [EB/OL] <http://t.csdnimg.cn/ZwWK2>

## Appendix

白马啸西风人名单											
李三	史仲俊	霍元龙	陈达海	上官虹	李文秀	丁同	苏鲁克	苏普	车尔库	阿曼	马家骏
碧血剑人名单											
张信	张朝唐	张康	王道	老王	杨鹏举	农夫	孙仲寿	徐达	朱安国	皇太极	李自成
刘芳亮	曹化淳	黄须人	白脸人	田见秀	崔秋山	倪浩	应松	罗大千	袁承志	农妇	祖大寿
哑巴	安大娘	安小慧	胡老三	袁夫人	穆人清	木桑道长	说不得	夏雪宜	张春九	明月	温仪
李岩	红娘子	龙德邻	沙老大	老李	荣彩	温正	明珠	小菊	温方山	温方悟	温方达
温方义	温方施	温南扬	崔希敏	温青青	上官	黄真	吕七	归辛树	冯难敌	马公子	杨景亭
郑起云	史秉光	史秉文	黎刚	闵子华	梅剑和	焦公礼	刘培生	万里风	罗立如	多尔袞	焦宛儿
吴平	洞玄	归二娘	归钟	洪胜海	朱元璋	沙通天	黄二毛子	程青竹	阿九	秦栋	谭文理
水鉴	孟伯飞	丁游	胡大哥	胡桂南	孟铮	孟铸	宝树	张若谷	冯同知	董开山	钱通四
彼得	若克琳	安剑清	侯飞文	范文程	宁完我	鲍承先	玉真子	童子	单铁生	沙广天	齐云
何铁手	何红药	潘秀达	岑其斯	水云道人	何惕守	刘宗敏	宋献策	陈圆圆	吴三桂	冯不破	冯不摧
石骏	韦小宝	教士	乾隆	常遇春	黄宗羲	乾隆皇帝					
飞狐外传人名单											
胡一刀	苗人凤	马春花	马行空	徐铮	商宝震	何思豪	阎基	田归农	胡夫人	补锅匠	南仁通
蒋调侯	南兰	苗若兰	商老太	童子	平四	胡斐	平阿四	王维扬	王剑英	王剑杰	老王
陈禹	殷仲翔	枯木	张总管	张召重	赵半山	空见	孙刚峰	吕小妹	徐天宏	骆冰	农夫
马真	乾隆皇帝	福康安	乾隆	胖商人	瘦商人	凤七	俞朝奉	蛇皮张	凤一鸣	邝宝官	中年武师
孙伏虎	尉迟连	杨宾	袁紫衣	刘鹤真	说不得	蓝秦	曹猛	崔百胜	胡大哥	风波恶	莫大
童怀道	张飞雄	程灵素	慕容景岳	姜铁山	薛鹊	王铁匠	姜小铁	田青文	上官	李自成	汪铁鹞
曾铁鸥	秦耐之	明月	无青子	大智禅师	褚轰	无尘道长	文泰来	聂钺	兆惠	姬晓峰	相国夫人
袁枚	明珠	蔡威	张九	任通武	吊死鬼	汤沛	海兰弼	安提督	黄希节	欧阳公政	西灵道人
文醉翁	周隆	郭玉堂	齐伯涛	陈高波	宗雄	哈赤大师	倪不大	倪不小	桑飞虹	无常鬼	常赫志
常伯志	上官铁生	刘之余	石双英	曹云奇	李廷豹	木文察	无尘道人	石万嗔	陈家洛	心砚	袁士霄
霍青桐	德布	陆菲青	余鱼同	李沅芷	王掌柜	司徒雷	乔峰	杨过	郭靖	令狐冲	0

连城诀人名单											
戚长发	卜垣	狄云	戚芳	言达平	万震山	周圻	吕通	吕威	鲁坤	万圭	孙均
吴坎	冯坦	沈城	桃红	丁典	臬道人	宝象	善勇	胜谛	凌霜华	凌退思	梅念笙
西施	玉如意	侯景	菊友	马大鸣	耿天霸	老王	渔人	阿三	枯木	农夫	水福
汪啸风	水笙	血刀老祖	明月	水岱	北四怪	刘乘风	花铁干	陆天抒	渡劫	哑巴	说不得
老不死	农妇	教士									

鹿鼎记人名单											
顾炎武	吕留良	黄宗羲	明月	吴之荣	鳌拜	庄允城	程维藩	本因	吴六奇	明珠	吴宝宇
陈近南	贾老六	韦春花	韦小宝	茅十八	王潭	吴大鹏	史松	吴三桂	徐达	常遇春	陈友谅
邓愈	汤和	赵钱孙	皇太后	平威	老吴	小玄子	温有道	温有方	童子	太皇太后	杰书
索额图	海大富	蕊初	玉林	说不得	张总管	崔瞎子	关安基	李力世	蔡德忠	方大洪	马超兴
林永超	李式开	姚必达	胡德第	云鹤	朱元璋	周颠	白寒松	白寒枫	樊纲	马博仁	姚春
雷一啸	王武通	徐天川	苏冈	卢一峰	老不死	风际中	吴应熊	多隆	神照上人	齐元凯	杨溢之
江百胜	吊死鬼	沐剑屏	方怡	瑞栋	劳太监	张康年	赵齐贤	刘一舟	钱老本	沐剑声	柳大洪
敖彪	吴立身	哑巴	李西华	柳燕	陶红英	大头鬼	邓炳春	章老三	双儿	庄夫人	渔人
于八	澄光	心溪	巴颜法師	皇甫阁	桑结	行颠	行痴	胖头陀	澄心	莫大	无根道人
张淡月	钟志灵	洪安通	许雪亭	殷锦	苏荃	陆高轩	瘦头陀	建宁公主	察尔珠	葛通	五符
元义方	富春	司徒伯雷	李自成	司徒鹤	曾柔	晦聪禅师	澄通	净济	净清	澄识	澄观
令狐冲	葛尔丹	昌齐	马宝	袁承志	呼巴音	郑克爽	农妇	施琅	冯锡范	九难	冯不破
冯难敌	阿九	穆人清	黄真	陈圆圆	玉真子	无色	阿三	大苦	祁清彪	夏国相	张康
西施	邝天雄	红娘子	胡逸之	胡大哥	林兴珠	木桑道长	南怀仁	汤若望	教士	杨光	黄甫
云素梅	苏菲亚	高里津	西奥图三世	彼得	皇太极	多尔袞	巴泰	卫周祚	对喀纳	米思翰	杜立德
图海道	上官	赵良栋	张勇	王进宝	孙思克	马真	马佑	慕天颜	张妈	归辛树	何铁手
何惕守	归钟	归二娘	阿济赤	彭参将	洪朝	路副将	图尔布青	华伯斯基	齐洛诺夫	佟国纲	马喇
阿尔尼	马齐	阿二	费要多罗	乾隆	菊芳	吕葆中	吕毅中	舒化龙			

三十三剑客图人名单											
范蠡	西施	王道	枯木	阿青	皇太后	苏辙	莫大	明月			

射雕英雄传人名单											
张十五	郭啸天	杨铁心	曲三	包惜弱	郭靖	杨康	李萍	段天德	江南七怪	盖运聪	姜文
完颜洪烈	焦木和尚	朱聪	韩宝驹	樵子	南希仁	张阿生	韩小莹	柯镇恶	全金发	桃红	上官
枯木	明月	铁木真	博尔术	赤老温	忽都虎	者勒米	窝阔台	术赤	察合台	拖雷	博尔忽
完颜洪熙	王罕	札木合	桑昆	都史	陈玄风	梅超风	黄药师	华筝	尹志平	谭处端	刘处玄
王处一	郝大通	孙不二	沙通天	吴青烈	钱青健	灵智上人	彭连虎	黄蓉	侯通海	梁子翁	穆念慈
王道	欧阳克	汤祖德	蓉儿	简管家	童子	说不得	陆乘风	何足道	莫大	靖哥哥	洪七公
周伯通	西施	范蠡	渔人	陆冠英	裘千仞	大苦	余兆兴	黎生	程瑶迦	老顽童	大头鬼
无常鬼	哑巴	明珠	傻姑	瘦丐	胖丐	简长老	鲁有脚	彭长老	梁长老	裘千丈	瑛姑
农夫	一灯大师	冯默风	天竺僧人	哑梢公	乔寨主	李志常	王志坦	祁志诚	赵志敬	耶律楚材	琴儿
杨过	吕文德	朱元璋	乾隆	蒙哥	忽必烈	阿里	谭公				

神雕侠侣人名单											
陆无双	雷猛	李莫愁	阿根	陆二娘	武三通	一灯大师	武三娘	农夫	武敦儒	武修文	黄蓉
明珠	柯镇恶	江南七怪	黄药师	洪七公	郭靖	郭芙	梅超风	双儿	陆乘风	郭大侠	杨过
杨康	王处一	穆念慈	欧阳克	蓉儿	赵钱孙	靖哥哥	朱聪	韩宝驹	南希仁	张阿生	陈玄风



鲁有脚	郭啸天	杨铁心	全金发	莫大	说不得	尹志平	灵智上人	郝大通	鹿清笃	赵志敬	小龙女
察合台	窝阔台	霍都	达尔巴	谭处端	程瑶迦	崔志方	孙婆婆	张志光	刘处玄	孙不二	林朝英
洪凌波	申志凡	西施	执法长老	教士	耶律晋	耶律楚材	耶律齐	耶律燕	完颜萍	藏边五丑	金轮法王
王十三	陆冠英	本相	渔人	朱子柳	梁长老	彭长老	点苍渔隐	傻姑	老顽童	周伯通	冯默风
忽必烈	潇湘子	尼摩星	马光佐	尹克西	子聪	瑛姑	何足道	公孙绿萼	童子	樊一翁	公孙止
裘千尺	裘千仞	裘千丈	柔儿	明月	张二叔	吕文德	鄂尔多	郭破虏	郭襄	拖雷	农妇
韩小莹	天竺僧	萨多	王道	祁志诚	李志常	王志坦	宋德方	大苦	瘦丐	胖丐	蒙哥
丁大全	陈大方	王惟忠	小王将军	宋五	大头鬼	煞神鬼	无常鬼	史伯威	史仲猛	史叔刚	史季强
丧门鬼	笑脸鬼	催命鬼	吊死鬼	俏鬼	讨债鬼	冲虚	小棒头	人厨子	百草仙	无色	圣因师太
张一氓	九死生	韩无垢	简长老	童大海	蓝天和	青灵子	沙通天	侯通海	彭连虎	阿里	杨夫人
觉远	无相禅师	觉远大师	张君宝								

书剑恩仇录人名单

李沅芷	李可秀	陆菲青	乾隆	焦文期	罗信	贝人龙	曾图南	赵半山	无常鬼	常赫志	常伯志
王维扬	阎世魁	阎世章	兆惠	童兆和	于万亭	戴永明	钱正伦	木卓伦	霍青桐	霍阿伊	说不得
孙老三	文泰来	骆冰	陈正德	关明梅	老吴	吴国栋	冯辉	蒋天寿	韩春霖	余鱼同	马真
张召重	老王	周仲英	无尘道长	宋善朋	孟健雄	周绮	周大奶奶	周英杰	瑞大林	万庆澜	成璜
言伯乾	卫春华	章进	徐天宏	无尘道人	陈家洛	杨成协	安健刚	石双英	上官	蒋四根	心砚
袁士霄	乾隆皇帝	韩文冲	喀丝丽	朱祖荫	曹能	平旺先	哑巴	唐六爷	曹司朋	梅良鸣	孙克通
王道	马善均	马大挺	尹章垓	白振	玉如意	范中恩	龙骏	褚圆	瑞芳	晴画	明月
农夫	樵子	汪浩天	马敬侠	福康安	凯别兴	宋五	皇太后	王剑英	王剑杰	明珠	袁枚
沈德潜	郑板桥	张安官	皇太极	顾金标	哈合台	彭三春	宋天保	覃天丞	上官毅山	孙大善	忽伦大虎
忽伦二虎	忽伦三虎	忽伦四虎	和尔大	德鄂	渡劫	阿里	桑拉巴	阿凡提	童子	胡老爷	周阿三
方有德	老不死	元悲	元痛	元伤	大痴	天虹禅师	大苦	大癫	天镜	迟玄	呼音克
教长	相国夫人										

天龙八部人名单

马五德	左子穆	段誉	龚光杰	渡劫	容子矩	天山童姥	司空玄	钟灵	说不得	阿胜	阿洪
高老者	明珠	逍遥子	岳老三	来福儿	钟万仇	段正淳	甘宝宝	瑞婆婆	皇太后	平婆婆	哑巴
葛光佩	木婉清	辛双清	叶二娘	云中鹤	莫大	朱丹臣	褚万里	傅思归	古笃诚	郁光标	吴光胜
明月	秦红棉	段正明	太皇太后	刀白凤	巴天石	一灯大师	黄眉和尚	华赫良	范骅	阿根	段延庆
过彦之	崔百泉	慧真	慧观	玄慈	慕容博	丁春秋	本因	本观	本相	本参	枯荣长老
鸠摩智	王道	何足道	阿碧	阿朱	幽草	慕容复	唐光雄	小翠	王语嫣	邓百川	严妈妈
渔人	姚伯当	司马林	诸保昆	姜师叔	孟师叔	包不同	乔峰	风波恶	矮老者	全冠清	张全祥
执法长老	传功长老	宋长老	奚长老	白世镜	李春来	项长老	刘竹庄	吴长风	公冶乾	陈孤雁	易大彪
徐长老	马夫人	谭公	谭婆	单叔山	单正	赵钱孙	单伯山	单小山	老不死	智光大师	单季山
赫连铁树	努儿海	阿二	农夫	农妇	玄苦	止清	玄寂	玄难	鲍千灵	祁六三	游驷
游驹	薛慕华	谭青	游坦之	朴者和尚	萧远山	阿紫	阮星竹	出尘子	摘星子	许卓诚	室里
白熊	耶律洪基	耶律涅鲁古	耶律重元	耶律莫哥	大苦	慧净	苏星河	天狼子	张中	虚竹	玄痛
慧方	西施	照虚	康广陵	范百龄	苟读	吴领军	冯阿三	李傀儡	石清露	桑土公	洞玄
木岛主	端木元	黎夫人	乌老大	不平道人	无崖子	李秋水	梦姑	余婆婆	石嫂	符敏仪	程青霜
云岛主	卓不凡	崔绿华	梅剑	兰剑	竹剑	菊剑	哈大霸	缘根	神山上	神音	哲罗星

人											
波罗星 少林老僧	天竺僧	玄生	天竺僧人	简捷	无色	梅剑和	闻先生	宗赞王子	贾老者	苏辙	穆贵妃

侠客行人名单

王掌柜	吴道通	周牧	矮老者	闵柔	李大元	安奉日	冯振武	元澄道人	耿万钟	谢烟客	王万仞
花万紫	石中玉	封万里	白万剑	柯万钧	丁不三	丁不四	孙万年	褚万春	白自在	老不死	大悲老人
黄面道人	丑脸汉子	司徒横	贝海石	石破天	米横野	侍剑	展飞	说不得	邱山风	陈冲之	明月
汪万翼	呼延万善	闻万夫	小翠	史小翠	张三丰	李万山	胡大哥	老李	王老六	尤得胜	天虚
冲虚	照虚	灵虚	高三娘子	范一飞	风良	吕正平	明珠	杨光	哑巴	石嫂	何足道
廖自砺	成自学	齐自勉	梁自进	木岛主	解文豹	西门观止	童子	郑光芝	梅芳姑	温仁厚	莫大
梅文馨											

笑傲江湖人名单

史镖头	林平之	郑镖头	白二	陈七	林震南	祝镖头	王元霸	于人豪	余沧海	说不得	农妇
刘正风	莫大	王二叔	陆大有	劳德诺	罗人杰	长青子	林远图	余人彦	梁发	施戴子	令狐冲
仪琳	田伯光	天松道人	向大年	米为义	天门道人	闻先生	明珠	渡劫	申人俊	木高峰	岳不群
曲非烟	陶钧	英白罗	岳灵珊	夏老拳师	史登达	万大平	左冷禅	曲洋	陆伯	刘菁	刘芹
舒奇	宁中则	风清扬	平一指	桃根仙	桃干仙	桃枝仙	桃叶仙	桃实仙	桃花仙	鲁连荣	哑巴
不戒和尚	英颢	王伯奋	王仲强	王家驹	王家骏	易师爷	绿竹翁	平夫人	祖千秋	老不死	白熊
蓝凤凰	农夫	游迅	严三星	西宝	玉灵道人	司马大	谭迪人	辛国梁	易国梓	向问天	天乙道人
乐厚	施令威	秃笔翁	童子	任我行	室里	鲍大楚	王诚	桑三娘	吴天德	仪和	仪清
郑萼	于嫂	钟镇	卜沉	沙天江	竹剑	易堂主	齐堂主	冲虚	震山子	何足道	无色
上官云	上官	杨莲亭	童百熊	玉玑子	玉磐子	玉音子	麻衣汉子	任盈盈	朱元璋	哑婆婆	周孤桐
吴柏英	玉钟子	樵子	张三丰	郭靖							

雪山飞狐人名单

乾隆	曹云奇	周云阳	殷吉	陶子安	田青文	田归农	陶百岁	熊元献	郑三娘	静智大师	刘元鹤
宝树	苗人凤	范帮主	赛总管	明珠	胡斐	童子	于管家	琴儿	苗若兰	李自成	吴三桂
李岩	胡一刀	胡大哥	王维扬	胡夫人	平阿四	阎基	刘宗敏	平四	乾隆皇帝	杜希孟	袁紫衣
程灵素											

倚天屠龙记人名单

小龙女	杨过	郭靖	黄蓉	无色	郭襄	觉远	觉远大师	张君宝	哑巴	无相禅师	郭芙
黄药师	鲁有脚	耶律齐	林朝英	瑛姑	武修文	一灯大师	老顽童	周伯通	完颜萍	裘千仞	江南七怪
潇湘子	尹克西	矮老者	郭大侠	方天劳	卫天望	何足道	潘天耕	杨夫人	天鸣方丈	苦智禅师	农妇
冲虚	张三丰	俞岱岩	蒙哥	德成	渔人	都大锦	马真	祝镖头	史镖头	张翠山	宋远桥
殷梨亭	莫声谷	张松溪	老王	空闻	圆音	圆业	圆心	农夫	殷素素	常金鹏	过三拳
麦少帮主	白龟寿	高则成	蒋涛	莫大	谢逊	元广波	麦鲸	空智	空性	殷天正	说不得
白熊	空见	成昆	张无忌	程坛主	封坛主	殷野王	西华子	卫四娘	静虚师太	唐文亮	史火龙
贺老三	何太冲	灭绝师太	泉建男	孟正鸿	本因	祁天彪	云鹤	宫九佳	殷无福	殷无禄	殷无寿
班淑娴	静玄师太	纪晓芙	贝锦仪	宗维侠	常敬之	常遇春	周芷若	胡青牛	丁敏君	彭莹玉	鲜于通
简捷	曾柔	杨不悔	薛公远	樵子	金花婆婆	王难姑	胡夫人	吊死鬼	杨逍	徐达	吴祯

花云	朱元璋	汤和	吴良	邓愈	苏习之	詹春	五姑	乔福	小凤	朱九真	朱子柳
武青婴	武三通	武敦儒	洪七公	卫璧	朱长龄	姚清泉	武烈	苏梦清	阳顶天	韦一笑	宋青书
明月	庄铮	吴劲草	周颠	张中	冷谦	小昭	胡青羊	高老者	渡劫	辛然	闻苍松
静空	唐洋	吴六破	明珠	赵一伤	钱二败	孙三毁	李四摧	周五输	郑七灭	王八衰	赵钱孙
赵敏	灵虚	阿二	阿三	刚相	老不死	方东白	摩诃巴思	鹿杖客	苦头陀	鹤笔翁	范遥
察罕特穆尔	王保保	乌旺阿善	拖雷	忽必烈	哈总管	殷离	季长老	郑长老	陈友谅	流云使	妙风使
辉月使	韩千叶	平等宝树王	俱明宝树王	齐心宝树王	宝树	常胜宝树王	王道	掌钵龙头	掌棒龙头	执法长老	传功长老
韩林儿	童子	小翠	小虹	小玲	史红石	刘敖	上官	秦老五	寿南山	杜百当	易三娘
马法通	邵鹤	邵燕	渡厄	渡难	郝密	卜泰	夏青	司徒千钟	叶长青	欧阳牧之	静迦
静慧	太虚子	静照									
鸳鸯刀人名单											
周威信	上官	逍遥子	常长风	花剑影	盖一鸣	任飞燕	林玉龙	萧半和	明珠	杨中慧	杨夫人
袁夫人	卓天雄	袁冠南	说不得	张妈	童子						
越女剑人名单											
范蠡	西施	郑旦	阿青								