# 基于2017年央视报道的国家热点挖掘

张驰俊

2018.08.24

**目录**

[基于2017年央视报道的国家热点挖掘 1](#_Toc522907352)

[1. 概述 2](#_Toc522907353)

[2. 报道关键词提取效果展示 3](#_Toc522907354)

[3. 文本聚类及相应词组匹对频数统计 4](#_Toc522907355)

[4. 层次主题提取与挖掘分析 6](#_Toc522907356)

[4.1 Topic 1 6](#_Toc522907357)

[4.2 Topic 7 8](#_Toc522907358)

[4.3 Topic 10 9](#_Toc522907359)

[4.4 Topic 5 10](#_Toc522907360)

[5. 关于2017年（“习近平”，“合作”）关键词词组的探究 11](#_Toc522907361)

## 1. 概述

新闻文本挖掘，作为自然语言处理的一个分支，旨在探究大量历史报道内容中的潜在价值信息，该过程不直接以得出结论为导向，而是在不断的数据整备中，寻找潜在于语料库之间的内在规律或规则，并进一步地利用挖掘到的知识指导后续数据分析，通过统计描述、话题聚类等手段以描述社会现象。

在李岱峰老师的指导下，**本人就2017年的“已播”央视报道（不包括不含正文描述的录像盒、录像磁等影音视频数据集以及非中文报导数据集）为语料库**，针对去重、数据清洗处理后2017年的**共86532篇新闻报导**，在归纳并整合停用词及报道常用词、BosonNLP、THUOCL开放中文词库的基础上，以jieba为分词工具，应用**LDA**(Latent Dirichlet Allocation)文档主题生成模型和**TF-IDF**算法，**提取每篇报道的关键词，并据此实现文本聚类**。同时，本研究在得到较为满意的聚类结果后，从中确定热点分析领域，继续提取相应的新闻报道数据集，有针对性地实现**层次主题提取以及重要关键词词组的统计分析**，从央视新闻的报道内容入手，**挖掘2017年国家层面的社会热点**。

完成如上系列任务所使用的代码、运行顺序及各自的功能如下：

1 2017\_data\_extractor.py 从内容混杂的xml新闻报导数据集中格式化抽取id, date, title, content，并保存为.data文件。

2 split\_file.py 将data文件进行拆分，提高后续的分词效率(多进程)。

3 cutsentence.py加载userdict.txt(用户词典)和stopwords.txt(停用词)，对数据集的每一篇报道进行分词（8G内存，3.6GHz虚拟机运行速度大概在10分钟左右）。

4 tfidf\_extractor.py 抽取每篇文章tf-idf数值最高的前n个关键词，作为一篇文章的标引

5 de\_duplication.py 将内容相同的文章进行排除，执行去重操作 。

6 topic\_month\_classifier.py 反复调参，寻找log likelihood绝对值更小、聚类后各主题下的内容更统一的结果。该代码将分类结果（字典）通过pickle模块保存下来。

7 2017\_kws\_pair\_mining.py 对topic\_month\_classifier.py聚类所得且值得关注的主题进行关键词挖掘，可以挖掘一个关键词、或者2/3/4个关键词词组匹对(keywords pair)的数量情况（包括随时间推移在各月/各星期的数量）

8 hierarchy\_topic\_extractor.py对topic\_month\_classifier.py聚类所得且值得关注的主题执行层次主题挖掘，观察该主题下又主要在讨论什么内容。同时还将以绘制折线图呈现方式统计该类主题下所有文章在一年间的报导分布情况。

总的来说，通过简单的参数调整，本项目代码可实现如下工作任务：

* 自定义待提取关键词量，对任一篇报道的主旨大义进行结构化抽取
* 自定义待划分类别量，以无监督方式对一批新闻报道进行主题聚类，同时统计各主题类目下新闻报道的数量
* 自定义关键词或关键词词组内容（如'一带一路' 或 ('一带一路', '合作', '高峰论坛')），统计任意指定时间段内包含该关键词或词组的报道数量变化
* 自定义关键词词组所包含的关键词量，展示每个聚类主题下出现频率最高的N组关键词词组

**本报告将略去代码要点详解、分词功能完善及模型调参等计算机层面的工作内容，重点展示所挖掘结论。**

## 2. 报道关键词提取效果展示

新闻报道的关键词提取采用TF-IDF算法，该算法本质上是一个统计方法，用以评估某一词语对于一个语料库中的其中一份文件的重要程度。简单来说，**TF-IDF算法认为：如果一个词语在某一篇报导中出现次数很多，而在语料库中各报导下出现的比重较低，那么该词语对于这一文章而言就很重要，有理由认定其能够在一定程度反映报导主要内容**。

任取其中一篇报道。报道正文内容如下：

68414669 2017-05-11T22:01:25（首发）（主标题）“一带一路”·合作共赢

高峰论坛即将到来，我们再通过漫画，了解下“一带一路”到底是什么，会给世界和我们带来什么改变。“一带一路”是“丝绸之路经济带”和“21世纪海上丝绸之路”的简称。2000年前开始，我们中国人就通过陆上和海上将丝绸、瓷器等产品卖到欧洲和沿线国家，形成了举世闻名的丝绸之路，今天，我们要用这个著名的历史符号引领助力世界经济复苏。丝绸之路经济带重点畅通中国经中亚、俄罗斯至欧洲；中国经中亚、西亚至波斯湾、地中海；中国至东南亚、南亚、印度洋。21世纪海上丝绸之路重点方向是从中国沿海港口过南海到印度洋，延伸至欧洲；从中国沿海港口过南海到南太平洋。除了沿途地区，全世界有意向合作的国家都可以参与。那么现在我们还是要卖丝绸和瓷器吗？当然不是，“一带一路”不是一个实体和机制，而是合作发展的理念和倡议，总结起来就是“五通三同”，“五通”是政策沟通、设施联通、贸易畅通、资金融通、民心相同。“三同”就是利益共同体、命运共同体和责任共同体，它们相辅相成，缺一不可。“一带一路”横跨欧亚，各国政策法律差异巨大，甚至技术、质量标准各异，建设中要打破这些隔阂进行合作，就要靠“政策沟通”求同存异，化解一切消极因素，为项目建设开启政策绿灯。“一带一路”上，不少伙伴经济条件不是很好，基础设施亟待建设。这就需要设施联通，俗话说要想富先修路。我们有着先进技术和经验，修桥、修路、建铁路、建港口都是一把好手，东西好，还不贵。完善交通网就好比打通任督二脉，经络畅通能让整个经济带健康有活力。贸易畅通是“一带一路”建设的重点内容，想要加大参与国的贸易量，让小伙伴们互相用上对方的好产品，就要打破壁垒，使贸易、投资和人员往来便利化，加强信息合作、海关、认证方面合作，拓宽贸易投资渠道。建设，贸易，小伙伴缺钱怎么办？这就说到了资金融通，我们创立了亚洲基础设施投资银行，给小伙伴贷款搞基建，以后慢慢还钱。修的过程中让当地老百姓有活儿干，修好还能惠及几代人。

提取TF-IDF值最高的前10个关键词（语料库为2017年去重处理后的86532篇新闻报导），提取结果如下：

丝绸之路 | 一带一路 | 贸易 | 畅通 | 经济带 | 小伙伴 | 沿海港口 | 合作 | 五通 | 建设

**依据关键词提取结果，可推知**本报导主要在论述丝绸之路、一带一路的内容，比较重点谈及到的是贸易事宜，关乎畅通沿途经济带，其中谈及到了“五通”的概念，通过与“小伙伴”合作的方式，从中国沿海港口出发，推动经济建设。这**与报道内容是相符的，这种关键词提取算法也是当今自然语言处理较为常用的方法之一**。

## 3. 文本聚类及相应词组匹对频数统计

应用主题模型(lda)对关键词提取后的86532条记录进行文本聚类。对主题(Topic)聚类数量进行多次调试，使得对数似然函数距离（误差）尽可能小、各主题最重要的前n个关键词尽可能反映同一主题。同时，对各主题下报道的主要内容分别进行词组匹对挖掘，选取频数（此频数指包含该匹对的报道量）较高的词组匹对，与主题关键词共同作为该聚类类别的主题概括参考。

**提取每个主题下TF-IDF值最高的前20个词语，同时分别统计各主题下的词组匹对频数，选出4到5个能够直观反映该聚类主题内容的匹对，结合如上两者推断2017年所有新闻报道所主要涉及的领域与内容**。结果整理如下表所示：（参数水平：n\_topics=20, n\_iter=2000 \_iter, random\_state=1）

表1 2017年86532条央视新闻报道的文本聚类及相应词组匹对频数统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主题 | 关键词 | 主要词组 频数 | 主题概括 | 报道数 | 比重 |
| \*Topic 0 | 游客 扶贫 生态 村民 首发 农业 地方 森林 农村 脱贫 景区 农民 种植 五年 内蒙古 火场 新区 宁夏 保护 土地 | ('百名', '红通', '追逃') 27  ('扑救', '森林', '火场') 23  ('扶贫', '脱贫', '贫困') 20  ('新区', '规划', '雄安') 17 | 涉及森林火灾、城乡规划、扶贫等内容 | 4167 | 4.82% |
| \*Topic 1 | 合作 中国 一带一路 习近平 十九大 李克强 会议 代表 报告 金砖国家 会见 国家 国际 世界 论坛 金砖 工作 新时代 委员 高峰论坛 | ('一带一路', '合作', '高峰论坛') 98  ('合作', '金砖', '金砖国家') 97  ('会晤', '厦门', '金砖国家') 79  ('一带一路', '中国', '合作') 57 | 金砖会议 及 “一带一路”相关报道 | 6362 | 7.35% |
| \*Topic 2 | 铁路 旅客 台风 高铁 列车 登陆 大桥 隧道 春运 出行 客流 公里 广东 假期 地铁 高速 旅游 天鸽 全国 沿海 | ('台风', '沿海', '登陆') 99  ('旅客', '春运', '铁路') 48  ('旅客', '铁路', '预计') 37  ('西安', '西成', '高铁') 35 | 交通出行、台风问题 | 4483 | 5.18% |
| \*Topic 3 | 救援 被困 地震 积水 洪水 水位 火灾 强降雨 山体 村民 受灾 现场 消防人员 事故 群众 转移 居民 消防官兵 供电 暴雨 | ('救援', '消防官兵', '被困') 31  ('垮塌', '山体', '茂县') 29  ('救援', '洪水', '被困') 25  ('供电', '用电', '部门') 22 | 救灾援助 | 5004 | 5.78% |
| \*Topic 4 | 孩子 医院 学生 患者 学校 医生 教育 医疗 手术 老人 儿童 老师 服务 幼儿园 家长 病人 药品 治疗 小学 培训 | ('医院船', '和平', '方舟') 30  ('家庭医生', '服务', '签约') 25  ('学生', '桃江县', '肺结核') 19  ('h7n9', '疫情', '病例') 18 | 医疗卫生 | 4207 | 4.86% |
| \*Topic 5 | 朝鲜 韩国 外交部 发言人 中方 华春莹 美国 半岛 陆慷 耿爽 萨德 朴槿惠 中国 总统 对话 部署 和平 局势 特朗普 文在寅 | ('华春莹','发言人','外交部') 127  ('萨德', '部署', '韩国') 96  ('朴槿惠', '检方', '韩国') 65  ('朝鲜', '特朗普', '美国') 46  ('决议', '半岛', '安理会') 45 | 对朝鲜半岛的中国外交方针 | 4499 | 5.20% |
| \*Topic 6 | 企业 涨幅 污染 收报 生产 跌幅 上涨 人民币 交易日 垃圾 食品 价格 报收 环保 产品 环境保护 治理 督查 指数 督察 | ('交易日', '收报', '跌幅') 70  ('交易日', '收报', '涨幅') 62  ('中间价', '人民币', '收报') 61  ('上证综指','收报','深证成指') 48 | 股市行情 | 3825 | 4.42% |
| \*Topic 7 | 增长 企业 同比 住房 投资 城市 服务 项目 市场 改革 增速 证监会 意见 租赁 全国 就业 资金 商务部 房地产 制造业 | ('同比', '增速', '增长') 59  ('增长', '外贸', '进出口') 47  ('制造业', '指数', '采购') 32  ('商品住宅', '城市', '涨幅') 31 | 市场经济 | 5869 | 6.78% |
| \*Topic 8 | 比赛 选手 中国队 决赛 考生 冠军 考试 大赛 高考 运动员 全运会 成绩 龙舟 金牌 冬奥会 世锦赛 参赛 项目 民法 赛事 | ('中华', '大赛', '龙舟') 32  ('考生', '考试', '高考') 29  ('总则', '民法', '草案') 25  ('世锦赛', '孙杨', '自由泳') 16 | 运动会、高考、民法相关 | 3224 | 3.73% |
| \*Topic 9 | 叙利亚 袭击 事件 伊拉克 组织 菲律宾 死亡 武装 土耳其 维和 中国 美国 英国 政府军 伊朗 地震 爆炸 墨西哥 俄罗斯 摩苏尔 | ('尔特', '杜特', '菲律宾') 61  ('地震', '测定', '震源') 51  ('叙利亚'，'代尔', '祖尔') 43  ('叙利亚', '政府军', '组织') 39 | 中东空袭、地震等带来的伤亡事件 | 3754 | 4.34% |
| \*Topic 10 | 台湾 文化 大陆 春晚 网友 两岸 电影 演员 作品 过年 演出 中国 游客 回家 传统 春节 新年 舞台 祖国 活动 | ('两岸','两岸关系','九二共识') 37  ('台湾', '国台办', '大陆') 24  ('春晚', '演员', '舞台') 22  ('台湾', '大陆', '游客') 21 | 两岸关系及节日庆典等喜庆相关 | 4979 | 5.75% |
| \*Topic 11 | 美国 特朗普 俄罗斯 日本 安倍 总统 普京 欧盟 法国 英国 卡塔尔 伊朗 欧洲 大选 关系 制裁 北约 美俄 候选人 德国 | ('总统', '特朗普', '美国') 133  ('俄罗斯', '特朗普', '美国') 93  ('欧盟', '英国', '谈判') 57  ('俄罗斯', '制裁', '美国') 55 | 国外政治 | 4538 | 5.24% |
| \*Topic 12 | 训练 导弹 航母 演习 部队 编队 美军 海军 演练 军演 作战 战机 日本 联合 俄罗斯 空军 目标 飞行 集团军 飞行员 | ('卡尔', '文森', '航母') 71  ('编队', '航母', '辽宁舰') 63  ('训练', '飞行', '飞行员') 52  ('八一', '表演队', '飞行') 34  ('中国空军', '训练', '远洋') 32 | 军事新闻 | 4459 | 5.15% |
| \*Topic 13 | 气温 高温 降温 降雪 冷空气 地区 降雨 南方 主播 东部 一带 北部 降水 局地 暴雨 强降雨 南部 预警 影响 内蒙古 | ('冷空气', '气温', '降温') 146  ('东北', '冷空气', '降温') 72  ('北方', '南方', '气温') 60  ('冷空气', '降温', '雨雪') 50 | 天气预报(主要是冷空气预警) | 4472 | 5.17% |
| \*Topic 14 | 卫星 发射 技术 火箭 天舟一号 海洋 载人 航天 潜水器 长征 中国 天然气 科技 深海 试验 太空 机器人 创新 运载火箭 地质 | ('天宫二号','天舟一号'，'对接') 58  ('发射', '火箭', '长征') 55  ('天宫二号','天舟一号','货运飞船') 51  ('勇士', '深海', '潜水器') 49  ('北斗', '卫星', '导航系统') 41 | 科技发展（航天、潜艇） | 3369 | 3.89% |
| \*Topic 15 | 大熊猫 游客 巡视 保护 野生动物 野生 旅游 自然保护区 动物园 动物 森林 熊猫 云南 湿地 中央 干部 候鸟 宝宝 村民 中央纪委 | ('中央', '巡视', '巡视组') 73  ('大熊猫', '成都', '繁育') 36  ('大熊猫', '放归', '野化') 33  ('大熊猫', '宝宝', '繁育') 29  ('八喜', '大熊猫', '映雪') 27 | 野生动物管理与保护 | 3441 | 3.98% |
| \*Topic 16 | 航班 飞机 机场 c919 班列 无人机 旅客 飞行 中欧 客机 渔民 首飞 试飞 渔船 航线 大型 航空 船员 厦门 首都机场 | ('c919', '试飞', '飞机') 57  ('中欧', '开行', '班列') 46  ('旅客', '滞留', '航班') 32  ('ag600', '水陆', '飞机') 31 | 航空事业 | 2821 | 3.26% |
| \*Topic 17 | 诈骗 公司 被告人 信息 网络 案件 犯罪 海关 走私 个人信息 平台 法院 警方 团伙 手机 公民 网站 传销 受害人 电信 | ('嫌疑人', '犯罪', '诈骗') 35  ('个人信息', '信息', '公民') 35  ('海关', '缉私局', '走私') 33  ('犯罪', '电信', '诈骗') 33 | 刑事案件（更集中于海关走私方面） | 4529 | 5.23% |
| \*Topic 18 | 交警 民警 驾驶员 司机 事故 高速 车辆 货车 轿车 驾驶 乘客 公交车 机动车 电动车 行人 面包车 停车 酒精 驾驶证 摩托车 | ('斑马线', '行人', '避让') 55  ('乘客', '公交车', '司机') 43  ('交警', '司机', '酒精') 42  ('民警', '驾驶员', '驾驶证') 32 | 交通事故 | 4326 | 5.00% |
| \*Topic 19 | 嫌疑人 民警 警方 犯罪 公安局 监控 孩子 分局 毒品 派出所 刑侦大队 受害人 女士 手机 小区 刑警大队 作案 男孩 两名 女孩 | ('嫌疑人', '犯罪', '警方') 142  ('作案', '嫌疑人', '犯罪') 43  ('受害人', '嫌疑人', '犯罪') 33  ('公安局','刑侦大队','民警') 27 | 民事/刑事案件 | 4204 | 4.86% |

**“主题概括”反映2017年央视新闻的报道方向，“报道比重”则可反映央视对于各主题内容的报道偏重情况**。根据上表的主题归纳情况与报道占比，**本报告接下来将重点关注2017年较为特殊的新闻主题**（如2003年的特殊新闻是非典，2008年是汶川大地震及北京奥运会），它们分别是Topic 1（金砖会议 及 “一带一路”相关报道，占总报道的7.35%），Topic 7（市场经济状况，占总报道的6.78%），Topic 10（两岸关系及节日庆典等喜庆相关报道，占总报道的5.75%），以及Topic 5（中国对于朝鲜半岛的外交方针，占总报道的5.20%）。

## 4. 层次主题提取与挖掘分析

层次主题提取的思路其实就是在如上的初步文本聚类的基础上，对目标主题下涵盖的所有报道再做一次文本聚类工作，对相关领域做更细致的了解。

（由于分析相似，此处不再逐一绘制表格）

### 4.1 Topic 1

Topic 1主要涉及2017年3月的**全国政协十二届五次会议**，2017年5月的**第一届“一带一路”国际合作高峰论坛**和2017年9月的**金砖国家领导人第九次会晤**相关的宣传与公告、会议报道与总结等一系列新闻报道。

相应报道数随月份推移所呈现的趋势变化如下所示：

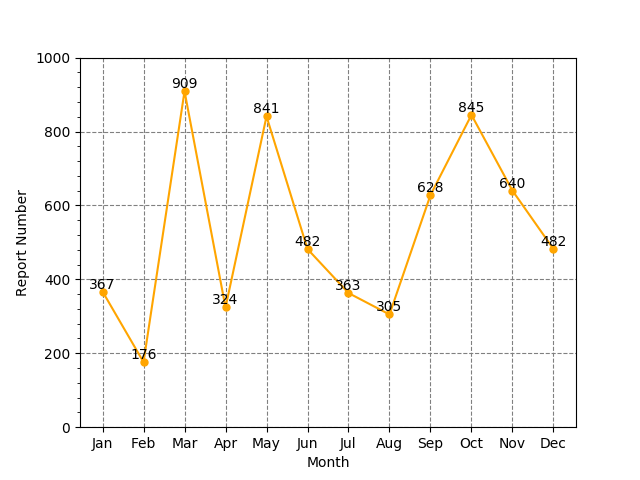


图1 隶属Topic1的央视报道量随月份推移的趋势变化图

经过资料整理，2017年1月，习近平总书记提出共同构建人类命运共同体；2017年3月，全国政协十二届五次会议召开，“一带一路”写入联合国决议、“一带一路”官网上线、“一带一路”成博鳌论坛热点等主要构成这段时间的909篇报道；2017年5月，中国与格鲁吉亚签署自贸协定、**“一带一路”国际合作高峰论坛举办**等，使该月的报道数也达到了841篇；2017年9月至11月期间，中国与黎巴嫩签署“一带一路”合作文件、“一带一路”写入党章、同时倡议与中东欧全体国家，其中已包括越南、摩洛哥、巴拿马，全覆盖式地签署“一带一路”合作文件。另一方面，2017年9月，金砖国家领导人第九次会晤在福建省厦门市顺利举办，拟推动团结协作，互利共赢。出于这些原因，使得Topic1的相关报道呈现如上折线图所示的数量变化。

列出几组能体现如上三大内容的关键词词组匹对：

* **43篇报道出现（“一带一路”，“合作”，“国际”，“高峰论坛”）；其中，2017年5月，同时出现“一带一路”、“合作”、“高峰论坛”这三个关键词的报道数达到96篇。**
* 42篇报道出现（“会晤”，“厦门”，“金砖国家”，“领导人”）
* 32篇报道出现（“习近平”，“思想”，“新时代”，“特色”）
* 27篇报道出现（“五次”，“会议”，“全国政协”，“十二届”）、

### 4.2 Topic 7

Topic 1主要涉及我国经济在制造业、对外贸易以及商品住宅的经济情况。

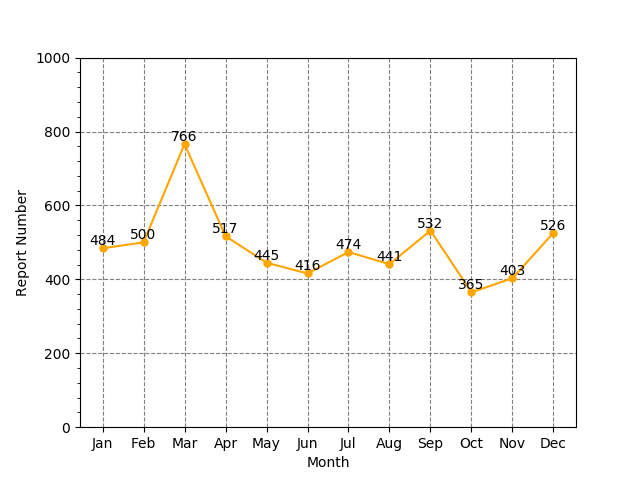


图2 隶属Topic7的央视报道量随月份推移的趋势变化图

结合层次聚类以及央视网的信息检索系统：

首先，**在该主题下最多篇文章中出现的关键词词组匹对是（“制造业”，“指数”，“经理”，“指数”）**，从国家统计局公布的2017年制造业与非制造业采购经理指数以及一些专家的解读来看，我国在**2017年的经济运行保持稳中趋升状态，结构持续优化，经济增长质量提升，需求小幅回升，以及经营活动增速加快**。

其次，是**外出打拼的年轻一代最为关注的房价问题，关键词词组匹对（“商品住宅”，“回落”，“城市”，“涨幅”）**。从国家统计局发布的2017年各月的70个大中城市住宅销售价格统计数据来看，**大中城市中一线城市房价环比整体呈持续下降态势，二三线城市涨幅也呈陆续回落或持平状态。房价高问题得到了一定的控制。**

最后，该主题还包括**进出口贸易的相关报道。以关键词“海关总署”、“进出口”、“外贸”、“增长”作为关键词进行检索**。整理结果显示：2017年我国对主要贸易伙伴进出口继续保持增长，欧盟、美国和东盟仍为我国前三大贸易伙伴，进出口分别增长15.5%、15.2%和16.6%，三者合计占我国进出口总值的41.8%。与此同时，“一带一路”成效不断显现也给我国外贸注入了新的活力，我国与“一带一路”沿线国家进出口继续保持快速增长势头，国际合作更加多元化。2017年，我国与“一带一路”沿线国家外贸增速为17.8%，高于同期我国外贸增速3.6个百分点。其中我国与俄罗斯、波兰、哈萨克斯坦等一带一路沿线国家进出口分别增长23.9%、23.4%和40.7%，均高于总体增幅。

在折线统计图中，该主题在2017年3月的相关报道较多，系为中国人民政治协商会议第十二届全国委员会做出了经济层面的相关体制改革与调整，除此之外，由于我国全年的经济状况都维持着稳定态势，整年关于经济情况的央视报道并未出现巨大波动。

### 4.3 Topic 10

Topic 10的重点内容是春节相关内容以及**大陆与台湾的两岸关系（“九二共识”）**。

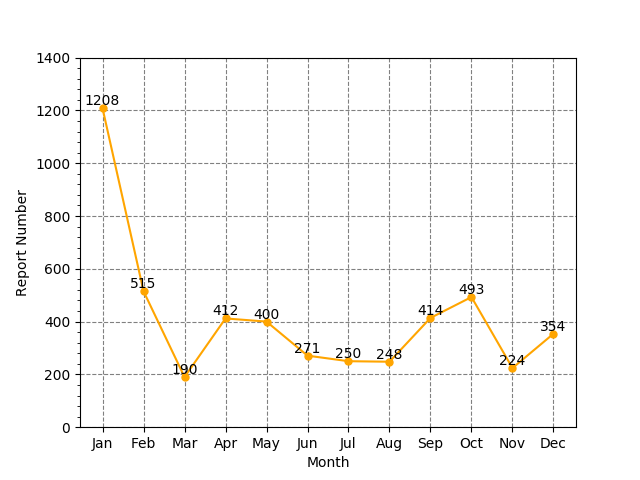


图3 隶属Topic10的央视报道量随月份推移的趋势变化图

2017年1月末2月初是春节假期，为此央视新闻做了一系列关于春晚、家族团圆和举国欢庆的一系列报道，因而该主题下一月与二月的报道数远高于其他月份。

除此之外，**该主题下余下的报道绝大多数都在探讨关于两岸关系的问题**。**2017年，涉及大陆与台湾两岸关系的关键词主要有“两岸关系”、“九二共识”、“国台办”、“马晓光”、“和平”。鉴于2016年5月20日以来，民进党当局拒不接受“九二共识”，却又强调台当局在那以来一直对大陆释放善意，马晓光代表国台办发布一系列声明，要求台当局领导人明确表明尊重“九二共识”的历史事实，认同两岸同属一个中国，也只有回到“九二共识”政治基础上，一切善意诚意都拿“九二共识”检验，两岸关系才能拨云见日。与此同时，大陆完全有信心并将继续推进与坚持“和平统一、一国两制”的方针，推动两岸关系和平发展，并将坚决挫败任何形式的“台独”分裂。两岸关系和平统一仍有不少阻碍，但从国台办保证将推出含金量更高的惠台措施也能够看出大陆一直在努力之中。**

### 4.4 Topic 5

Topic 5 **主要讨论朝鲜半岛的军事形势问题**。这其中的主要内容是萨德部署与中、美、韩、朝的外交表态。这期间代表我国外交部的发言人主要是陆慷、耿爽和华春莹。

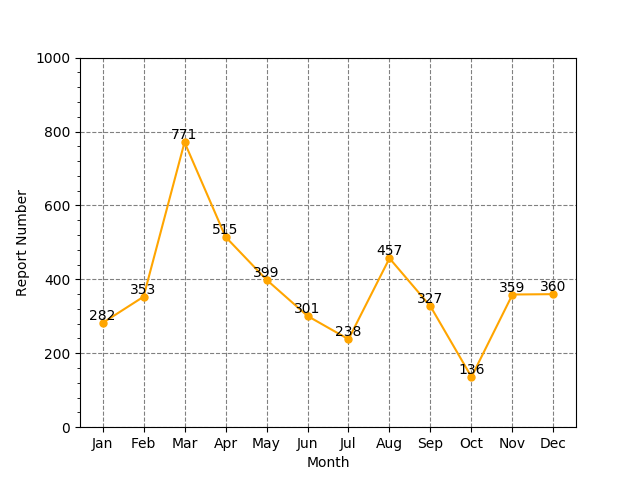


图4 隶属Topic5的央视报道量随月份推移的趋势变化图

据百度百科，2017年2月27日，乐天集团董事局决定，为部署萨德供地。3月7日上午，韩国国防部发布消息，萨德系统的部分装备，于2017年3月6日通过军用运输机运抵驻韩美军乌山空军基地。 2017年4月26日，萨德反导系统的2辆发射车、X波段雷达开始在韩国投入运行，2017年5月30日，韩国总统文在寅听取报告时得知，除部署与庆尚北道星州的2辆“萨德”拦截弹发射车外，其后又有4辆发射车被暗中运入韩国境内，文在寅要求彻查4辆发射车入境原委。**关于这一系列事件，央视在2月到5月之间对萨德给予了很高的关注**。

**然后是关于2017年8月份的报道，这段时间主要是联合国安理会于8月5日一致通过美国起草的2371号决议，决定对朝鲜实施制裁**。该决议将加强对朝鲜货物出口的制裁并限制对该国的投资，要求朝鲜停止所有弹道导弹发射活动与核试验。安理会要求恢复六方会谈，以和平方式确保朝鲜半岛无核化，以及美朝双方尊重对方主权并和睦共处。**美国方面，其国务卿蒂勒森宣传美国不是朝鲜的敌人，表示仍愿意与朝鲜对话，但由于朝鲜对美“无条件对话”无感，特朗普政府放狠话，并于11月将朝鲜再度列入支持恐怖主义国家名单。**

## 5. 关于2017年（“习近平”，“合作”）关键词词组的探究

在2017年一年里，包含该词组的报导数随时间推移的变化趋势图如下所示：

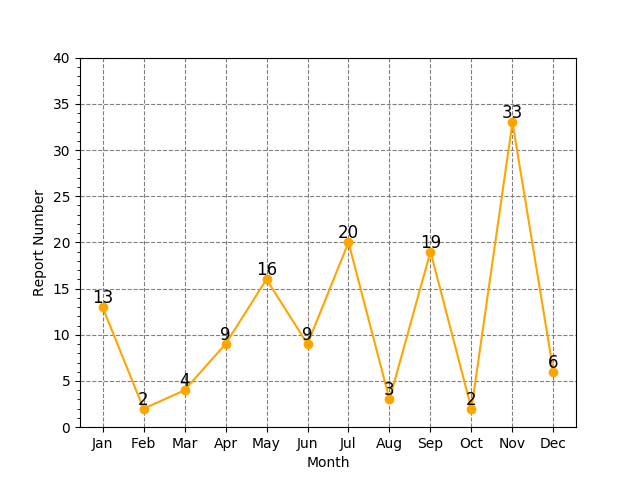


图5 2017年包含（“习近平”，“合作”）词组的报道量随月份推移的趋势变化图

上图显示了几个峰值，参考<http://fangtan.people.com.cn/GB/415888/index.html#m07> 人民网对习近平总书记于2017年的外交出访资料，可推知这些数值分别是如下事件的反映：

2017年1月：对瑞士进行国事访问、出席**世界经济论坛2017年年会**并访问瑞士国际组织。

2017年5月：第一届**“一带一路”国际合作高峰论坛**。

2017年7月：习近平出席《深化粤港澳合作 推进大湾区建设框架协议》签署仪式；对俄罗斯、德国进行国事访问并**出席二十国集团领导人第十二次峰会**。

2017年9月：**金砖国家领导人第九次会晤**。

2017年11月：对习近平主席在**中国共产党第十九次全国代表大会上的报告作总结**(<http://www.gov.cn/zhuanti/2017-11/17/content_5240304.htm>)；另外，习近平主席出席**亚太经合组织第二十五次领导人非正式会议**并对越南、老挝进行国事访问。