聊天记录分析之变量说明

本次分析报告所使用的数据均来自即时通讯软件QQ，共包含来自三个不同类型群组（朋友群、大学同学群和校园交易群）的近两万行文本数据。由于数据格式较为规范，可以提取出发言时间、发言用户和聊天内容三类变量。经过对聊天内容初步探索后，本文根据内容类型将其进一步划分为图片/表情、应用外链接、群应用和纯文本四个变量。具体变量说明表如表1所示。

表1 “聊天记录”数据变量说明表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **变量名** | | **变量类型** | **取值范围** | **备注** |
| 时间 | | 连续变量 | 2017/10/09-2017/10/24 | 包含日期、时间 |
|
| 用户名 | | 文本变量 | 包含昵称、QQ号 | 由于涉及用户隐私，在本文分析中将  昵称与QQ号换为游戏或漫画中的人物名称 |
|
| 聊天内容 | 图片/表情 | 文本变量 | 三种格式 【详见备注】 | 图片、表情在导出的聊天数据中表现格式有以下三种： [图片] [表情] /托腮 |
| 应用外链接 | 文本变量 | 包括网页分享和其他应用数据分享（如通过音乐软件分享的音乐） | 如： http://s.kugou.com/song.html?id=5DXAmdarAV2 |
| 群应用 | 文本变量 | 有固定格式 【详见备注】 | [群签到] [QQ红包] [群礼物]等 |
| 纯文本 | 文本变量 | —— —— | —— —— |