本次例会讨论了对每段报警信息是给出单一的类型还是给出标签？初步考虑以聚类产生的层次性类型为主，对某些特别的报警信息可以再加上副标签

**预处理组：**

孙浩然尝试了地名的命名实体识别，建议尝试获取更小的地名

杨迈提出：

采用余弦距离比欧氏距离效果好

同时采用多种聚类方法，提取其公共部分并增加其重复关键词权值。并减少部分其中的低频词汇的权值。然后再次尝试聚类

建议尝试各种不同的构建同义词表的方法

建议尝试各种不同的构建固定词组表（自定义词典）的方法

**聚类和分类组：**

建议利用聚类尝试自动构建分类树（并对树的每个节点加关键词标签，对叶子节点加建议的处理方法）的各种方法

**以前讨论的总结：**

尽量减少人工介入而自动挖掘，例如自动产生关键词表，停用词表，同义词表。可以考虑使用聚类结果作为评价标准