**西南交通大学**

**期中课程作业**

课程名称：拨云见物：从数据、方案到产品

课程代码：MYSC002615

2020-2021第1学期

## 新冠肺炎疫情期间各地口罩的紧缺情况——以成都为例

## 1 研究目的

2019年12月以来，新型冠状病毒肺炎逐渐肆虐全球，截至2020年5月10日，全球累计确诊人数已逾400万。

我们小组希望可以通过机器学习、构建神经网络模型等方式，构建一个对某地区是否存在口罩紧缺现象进行判断、从而达到尽早调配的目的的模型，推演出成都地区对口罩的紧缺情况。

## 2 数据来源

企业信息来源：企查查、百度信誉

人口数据：国家统计局、国家住建部

疫情数据：国家卫健委、各地卫健委、丁香园

百度热度指数：百度

## 3 研究结果

众多事实表明，疫情期间做好个人防护可以极大地降低感染风险，保护自己。我们小组注意到，在疫情刚刚爆发之际，由于个人防护用品的缺失，个人防护用品一时间供不应求，社会上出现了“口罩热”，医用口罩供不应求，成为了人人争抢的香馍馍。然而，结合我们组员的个人经历，发现并非所有地区都缺乏口罩，更多的是口罩分布的不均衡导致的短缺。而疫情愈演愈烈后，该现象也越发明显。

经过我们小组的分析，初步认为：在疫情期间，需要判断出各地口罩的紧缺状况，并及时乃至提前调配。这是防疫工作的重要一环。

口罩短缺的情况，会随着城市的具体情况而改变。经过我们小组的激烈讨论，决定从生产端与消费端两方面着手，考虑不同因素对口罩需求量的影响。

生产端方面，主要影响因素有两个：上游原材料供应商，如：熔喷布厂、无纺布厂、聚乙烯厂、纺织机械厂等，和下游生产商，如：口罩生产厂、医疗设备厂商等。生产商的分布与生产能力影响了它周边地区的口罩紧缺程度，而上游原材料商的分布情况则直接影响了生产商的生产能力。如广州地区，依靠附近的数家大型口罩厂，以及熔喷布等原材料厂商，迅速生产出大量口罩，不仅能够解决本地需求，还可以调配至其他相邻的紧缺地区。根据我们从企查查和百度信誉上调查得到的全国医疗口罩厂商的位置分布，得到图1、2：<http://mqn5xt.coding-pages.com/>

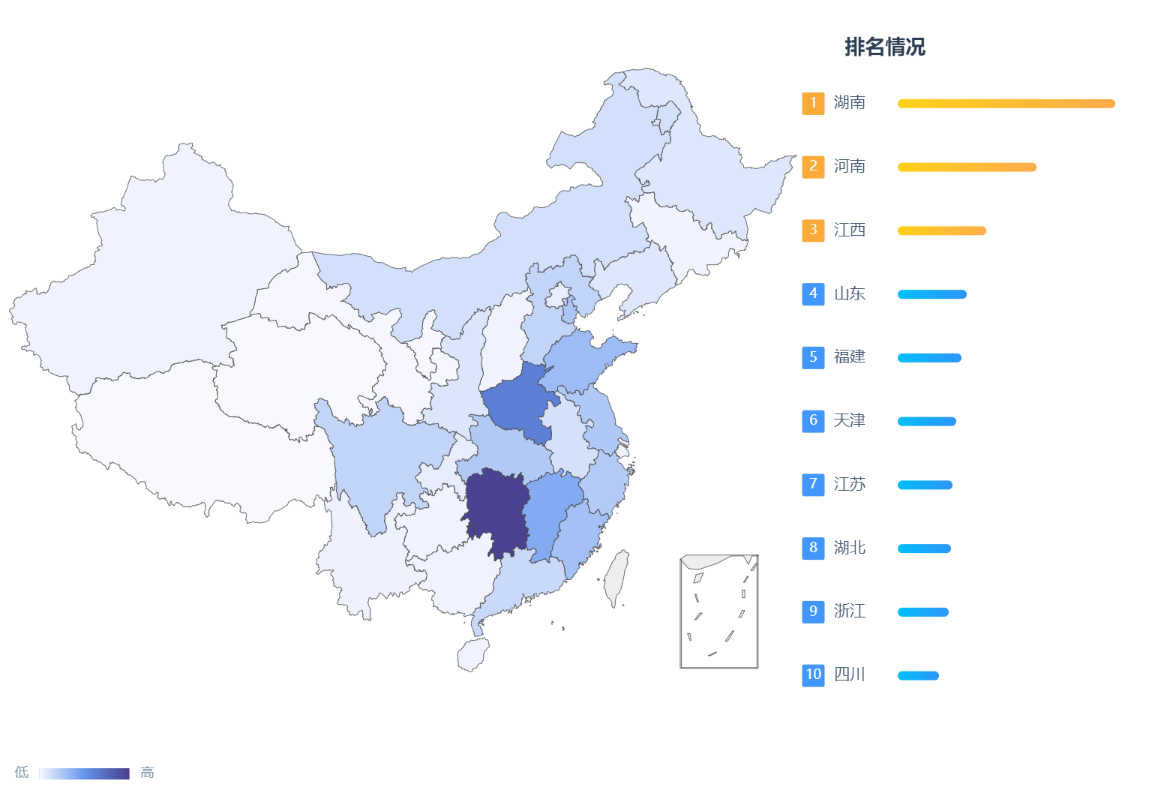


图1

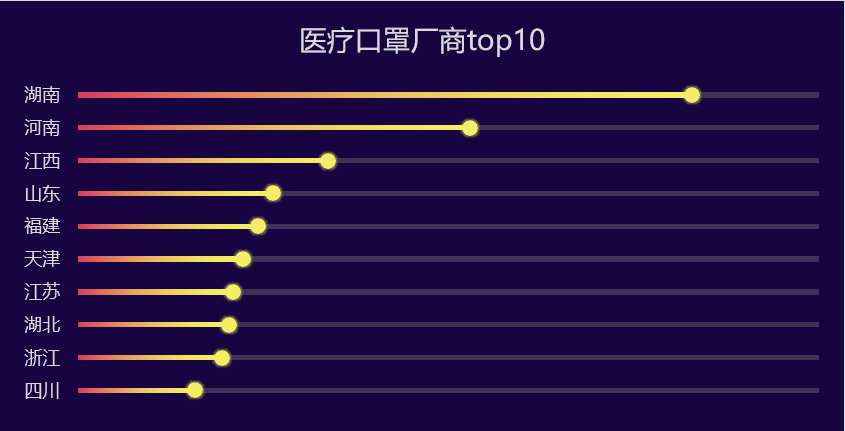


图2

消费端方面，人口数量与人口密度的影响极大，而疫情中感染人数的影响也不容忽视。此外，从关键词“口罩”的百度指数也能间接反映群众对口罩的需求程度。较大的人口密度十分有利于疾病的传播，而感染人数等则体现了未来该地区的感染潜力。

最后，我们对口罩短缺期的判断的准确与否，可以由该地区随后一段时间增加的确诊人数加以验证。如果口罩短缺，确诊人数应当会出现较大的增加；反之亦然。

我们最终从小组讨论的20个字段中确定出15个有效字段，根据已有的12883443条数据，选择SGD算法构建神经网络模型（源代码与数据集见：<https://github.com/kakasearch/data-course-publish>）。通过代入已有的大量数据，使用训练集进行训练，最终训练完成的结果如图3所示，正确率已经超过98%。

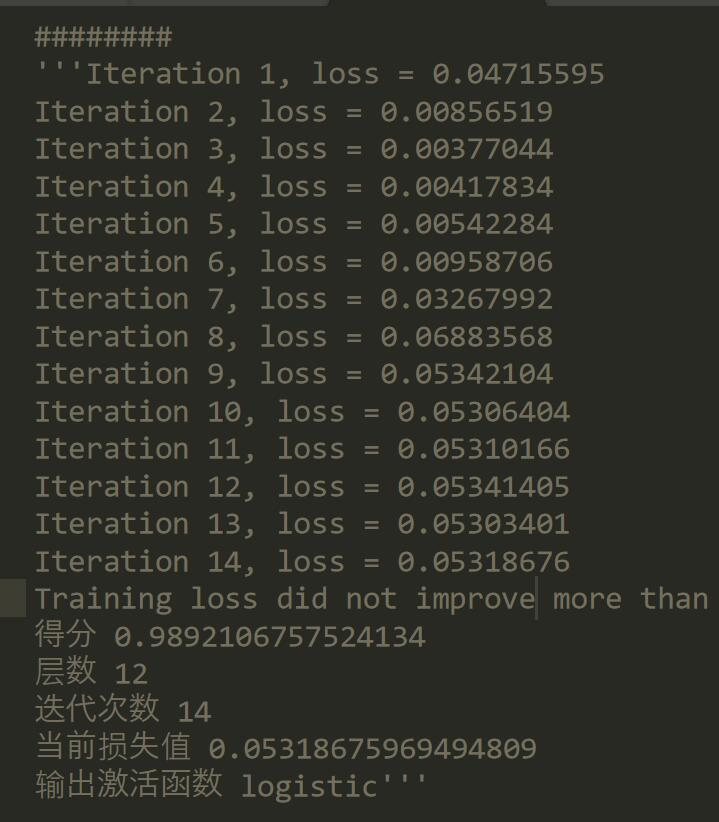


图3

接着我们使用训练完成的模型，通过代入较早阶段的数据，对成都地区的口罩紧缺状况做出预测。成都地区相关数据如图4、5、6，预测结果与实际结果对比如图7：

图4

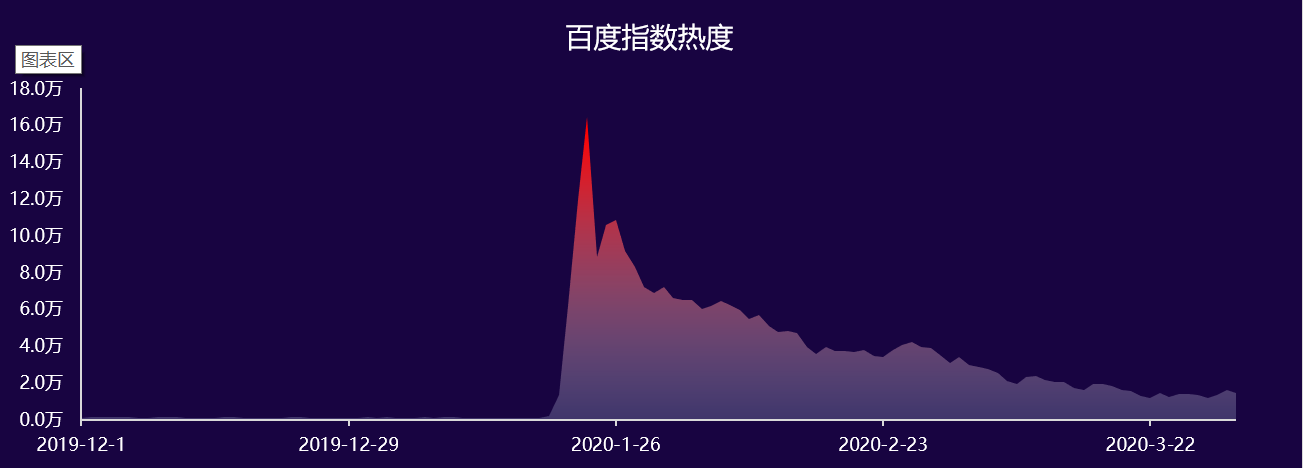


图5

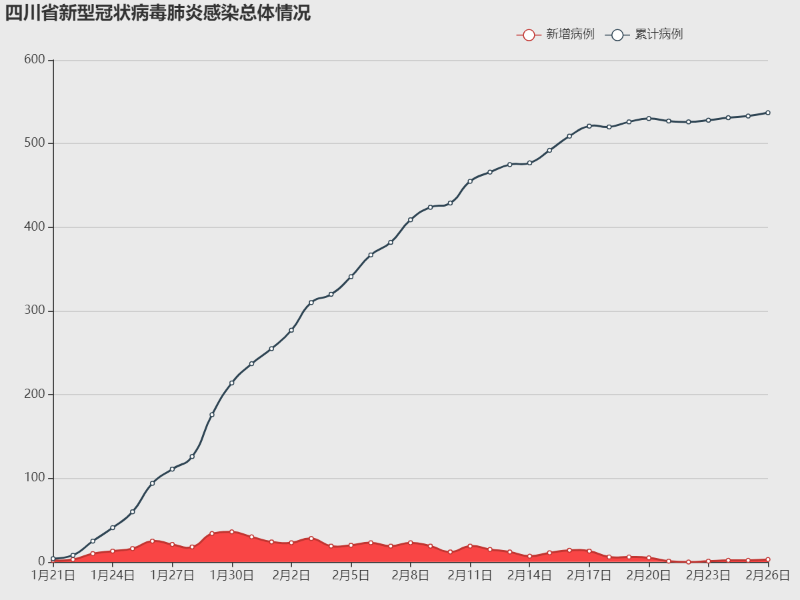


图6、四川省新冠病毒肺炎感染情况：<http://6s0dpr.coding-pages.com/>

图7 预测结果与实际情况对比

通过代入较早阶段的数据，我们推断出成都的两个口罩紧缺期。根据新闻报道，在1月22日左右，成都确诊病例开始增多，成都“口罩”的百度指数随即出现了极大幅度的上涨。考虑到成都作为人口破千万的特大城市，人口密度大，交通发达，极易出现大量感染者，口罩需求应该极大，极可能出现口罩的短缺。医学研究表明，新冠肺炎的潜伏期约14天，可以看到，在我们预测的口罩短缺期后14天左右，成都市确诊人数迅速增加。而作为西部重要城市，短缺期后成都的口罩资源能够迅速得到调配，事实上确诊人数增长速度也确实随之减缓。

可以看到，我们的模型对成都口罩短缺期的判断与实际情况是基本相符的。

## 4 主要结论

1、我们组的模型对成都市口罩紧缺期的判定是2020/1/23和2020/1/28-2020/2/1，所预测的结果与事实相比较，是基本一致的。

2、我们小组的模型考虑变量众多，但经过模型整合，所得到的结果与事实基本相符，并且在对成都市具体情况的判定上得到了验证。事实证明我们的模型的建立是比较成功的，对各个变量的权重的把握也比较准确。

3、事实证明，在发生重大疫情期间，合理高效的医疗物资调配方案极其重要。尤其是对于大型乃至于特大型城市，当地物资供需关系的变化幅度极大，考虑到物资调配的延后性，亟需对物资建立合理有效的调配方案。否则一旦延误，会出现大量不必要的伤亡。

4、不仅仅是疫情期间，在发生重大灾害时，合理高效的物资调配方式往往可以极大地减小损失，而如果能够做到对物资紧缺点的准确判断，带来的好处不可估量。

5、此次疫情状况存在极大的特殊性。一是私人的城际交通近乎停滞，快递等行业优先运送医疗物资，这也是我们小组讨论后放弃网购平台数据的原因之一（另一原因是网购平台存在刷单、隐私保护等问题，数据不够准确）；二是波及范围极大，一般情况下的灾害很少会波及全国乃至全世界，所以一般情况下可以全国各处集体支援受灾地而不必过分考虑物资在全国范围内的调配。

6、由于已有各地疫情数据，所以我们采用的是已知的感染人数。但是要想在更普遍的情况下进行预测，还要结合流行病学模型（如：SEIR、SEIRS模型等），并考虑网购等情况。这已经超出我们组的能力范畴，还需要更多的努力才能实现。