



2018年美赛B题论文点评

《美国数学建模竞赛》

完整课程请长按下方二维码





评判初定：论文摘要



2018 MCM Problem B: How Many Languages? 多少种语言?

- 背景：目前地球上约有6,900种语言。 普通话（包括标准中文），西班牙文，英文，印度文，阿拉伯文，孟加拉文，葡萄牙文，俄文，旁遮普文和日文，大约有一半的世界人口是以下十种语言中的一种。 但是，全世界的许多人口也会说第二种语言。 在考虑特定语言（母语人士，第二或第三等语言发言者）的总发言人数时，语言及其顺序会从所提供的母语列表中变化。
- 一种语言的说话人总数可能随着时间的推移而增加或减少，原因是各种影响包括但不限于一国政府使用和/或推广的语言，语言，用于学校，社会压力，文化团体的移民和同化，以及与讲其他语言的国家的移民。 而且，在我们这个全球化的，相互联系的世界里，还有另外的因素让地理上遥远的语言能够互动。 这些因素包括国际商业关系，增加的全球旅游业，电子通讯和社交媒体的使用，以及利用技术来协助快速和简单的语言翻译。



- 问题：一家在美国纽约市和中国上海设有办事处的大型跨国服务公司正在不断拓展，成为真正的国际化公司。该公司正在调查开设更多的国际办事处，并希望每个办事处的员工都能用英语和一种或多种其他语言进行讲话。该公司的首席运营官已经聘请了你的团队来调查全球语言的发展趋势以及新办公室的位置选择。
- 第一部分：
 - A. 考虑上述背景段落中描述的影响和因素，以及您的团队可能识别的其他因素。根据预测的趋势，以及一些或所有这些影响和因素，建立各种语言使用者随时间分布的模型。



- **B. 使用你的模型来预测在未来50年以英语为母语者和以英语为母语者的人数会发生什么变化。你认为目前排名前十的语言(母语或总母语)会被另一种语言取代吗?并做解释说明。**
- **C. 鉴于未来50年预测的全球人口和人口迁移模式, 这些语言的地理分布在同一时期是否会发生变化? 如果是这样, 请描述这个变化。**



- 第二部分：
- A. 根据您在第一部分中的模型，并假设您的客户公司想要开设六个新的国际办事处，那么您可以将这些办事处设在哪里？这些办事处使用什么语言？你的建议是短期的还是长期的？解释你的选择。
- B. 考虑到全球通信的变化，为了节省客户公司的资源，您是否可以建议公司开设少于六个国际办事处？说明你需要哪些额外的信息，并描述你将如何分析这个选项，以便为你的客户提供建议。



- 第三部分：
- 向服务公司的首席运营官写一个1-2页的备忘录，总结你的结果和建议。
- 注意：在您的分析中，忽略不可预测的或高影响的低概率事件，例如小行星碰撞，随着时间的推移会导致演化趋势发生灾难性的跳跃，并可能使所有语言都灭绝。



评判初定：论文摘要



2018 MCM Problem B: How Many Languages? 多少种语言?

- Globalization has promoted the spread and development of languages. How to accurately predict the distribution of each language has important research significance.
- 全球化促进了语言的传播和发展。如何准确预测每种语言的分布具有重要的研究意义。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- For Part1.A, conduct a correlation analysis on the factors(economy, etc.) that affect language development. Filter out five indicators. Then, we use partial least-squares regression to convert the linguistic index (economic, etc) into Leslie's traditional parameters (birth rate, etc). Improved Leslie model, called: Social-Leslie model.
- 第一部分。A、对影响语言发展的因素(经济等)进行相关分析。筛选出5个指标。然后，我们使用偏最小二乘回归将语言指标(经济等)转换成Leslie的传统参数(出生率等)。改进的Leslie模型称为: Social-Leslie模型。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- Amalgamating the countries where both the first language and the second language are the same. What's more, based on the country's influence in the world, the small and densely distribute countries are viewed as a whole. Thus, divide the six continents into 15 regions. We set up a corresponding Leslie matrix for all regions. In this way, the distribution of various language speakers can be modeled over time (see Figure 3,4).
- 将第一语言和第二语言相同的国家合并。更重要的是，根据这个国家在世界上的影响力，把人口密集的小国视为一个整体，如此，就把六大洲分成了15个区域。我们建立了所有区域的Leslie矩阵。通过这种方式，可以对不同语言使用者的分布按时间进行建模(如图3、4)。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- For Part1.B, based on the Social-Leslie model. In order to predict the number of non-native speakers, Mendelian law of genetics is introduced innovatively to divide the different language populations through significant and recessive genes. We predict the current top 10 languages and other languages in the world. Get the trend of different languages in the next 50 years. The result shows that French overtook Japanese in 2053 and entered the top ten. We think the main reason is that French is one of the five official UN languages and be used in many countries.
- 第一部分。B，基于Social-Leslie模型。为了预测非母语者的数量，创新性地引入了孟德尔遗传学定律，通过有效的隐性基因来划分不同的语言群体。我们预测了目前世界上排名前十的语言和其他语言。得到未来50年不同语言的发展趋势。结果显示，法语将在2053年超过日语，而进入前十。我们认为主要原因是法语是联合国五大官方语言之一，在许多国家都被使用。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- For Part1.C, considering the first language of the child born after immigration is different from the parent. According to immigration data in recent years, the fitting function is established. Based on this, we expand the Social-Leslie model and call it the improved Social-Leslie model. Get the geographical distribution of languages in the next 50 years. The results show that the distribution of Putonghua in West Asia has obviously increased, and the number of people using English in Southeast Asia has rapidly increased. Other languages also have an expanding trend, showing that international exchanges have been strengthened.
- 第一部分。C，考虑到移民后出生的孩子的第一语言与父母不同，根据近年来的移民数据，建立了拟合函数。在此基础上，我们扩展了Social-Leslie模型，并称之为改进的Social-Leslie模型。得到未来50年语言的地理分布。结果表明，西亚地区普通话分布明显增加，东南亚地区使用英语的人数迅速增加。其他语言也有扩大的趋势，表明国际交流得到加强。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- For Part 2.A, we evaluate the office's competitive edge at different addresses firstly. Then model is built by Data Envelopment Analysis to calculate input-output. Taking the economic level(cultural level, etc)as input indicators, the number of native speakers and second language population as output indicators. Analyze the long-term and short-term development separately and calculate the scale income of the office. Finally, according to the current situation of development, we get two groups of site selection plan (See Table 10,11).
- 第二部分。A，我们评估办事处建在不同地址的竞争优势，然后通过数据包络分析建立模型，计算输入输出。以经济水平(文化水平等)为输入指标，以母语人数和第二语言人口为输出指标，分别分析长期和短期的发展，计算出办事处的规模收益。最后，根据发展现状，我们得到两组选址方案(见表10、11)。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- For Part 2.B, we analyze the geographical location of the office, social position. Then according to the characteristics of the data, set up an optimization strategy of the number of offices and regional distance, the level of the communications industry, network security and stability, and economic level. The final result shows that the number of offices should be reduced to four: Tswana, Sydney, Buenos Aires, Berlin.
- 第二部分。B，我们分析办事处的地理位置、社会地位，然后根据数据的特点，建立了办事处数量和区域距离、通信行业水平、网络安全稳定、经济水平的优化策略。最后的结果表明，办事处的数目应该减少到四个：茨瓦纳、悉尼、布宜诺斯艾利斯、柏林。



2018 MCM ProblemB: How Many Languages: 有多少种语言

- Last but not least, we evaluate and promote the advantages and disadvantages of the model.
- 最后，我们做了该模型的优缺点评价和推广应用。



分数核定：论文内容



- Task1
- 首先选择指标和研究领域。然后收集数据并进行预处理。孟德尔遗传定律被创新地引入，通过重要的隐性基因来划分不同的语言群体。使用偏最小二乘回归将语言指数（经济等）转换为莱斯利的传统参数（出生率等）。改良的Leslie模型用来预测各种语言使用者的分布随时间的变化。考虑到移民后出生的孩子的第一语言的影响与父母不同。通过改进的Social-Leslie模型获得未来50年语言的地理分布。
- 步骤：
 - 1区域划分
 - 2指标选取
 - 3相关分析（指标筛选、语言水平选择）
 - 4数据收集和处理（异常值处理）
 - 5改进的社会-莱斯利模型



- 由于发言者人数的增加与当地人口增长之间的直接相关性，我们将发言者人数的增加转化为人口增长。我们考虑移民对后代语言的影响。（1）一个P国的家庭可以说不止一种语言，他们的孩子必须会说一种语言P。（2）一个国家的家庭移民到国家Q，他们的孩子应该是母语为Q的人。因此，我们将总人口分为母语人士和非母语人士；对于非母语人士，以每个大陆上相应语言的比例为概率，预测研究该语言的后代人数。
- Step1 社会-莱斯利模型的建立
- Step2 介绍所需参数——男女初生比例、各年龄段妇女的生存比例、男女比例
- Step3 改进社会-莱斯利模型的约束条件
- Step4 基于孟德尔遗传学定律的L2预测
- Step5 改进社会-莱斯利模型的建立——建立人口迁移拟合函数、建立模型



- 结论：前25年，普通话人口增长缓慢，25年后缓慢下降。日本的增长非常缓慢，其他语言也在上升。基于社会-莱斯利模型，我们也预测了世界上的其他语言。获取未来50年以该语言为母语的人数和总人数。比较显示法国人在2053年超过了日本人，进入了前十名。我们认为主要原因是：（1）法语是联合国五种正式语文之一。（2）有更多的法语国家。（3）法语和英语有“亲和力”，所以更容易学习。



- 考虑到人口迁移对语言发展的以下影响：（1）不同地区人口基数的变化；（2）促进语言交流；（3）移民后出生的子女与父母之间第一语言差异的影响。据此，得到了改进后的“社会列”模型，得到了未来50年语言的地理分布几年了。
- 图4，左边是当前英语的分布，右边是50年后英语的分布。相比之下，我们可以看到英语在世界范围内的传播增加了。同样，用其他语言得到结果。结果显示：西亚地区普通话分布明显增加，东南亚地区使用英语的人数迅速增加了。其他语言也有扩大的趋势，表明国际交流得到了加强。但是，日语的分布不明显。



- Task2

- 在第一部分A的基础上，建立了区域内各语言发展的评价体系，以及抽象的投入产出模型。评价指标包括本区域的经济水平，文化水平、社会稳定、信息水平发展。因为母语人士人数的增长率是一个地区发展的重要指标，所以第二语言的增长率可以反映该地区的发展。因此，我们制定了第一语言增长率、第二语言增长率的产出指标。国内生产总值投入指标学生在高等教育中的比例，就业率，互联网用户的比例。



- 该系统由多输入和多输出的复杂系统组成. . 为了更好地评价各地区的发展现状，使选择有利于未来的发展办公室的装饰。我们首先使用数据包络分析来避免每个指标的权重不准确。第二，系统中的每个输入都与一个或多个输出有关。这种关系的解决是非常复杂的。数据包络分析避免了决定这种关系的表达式。
- 综上，我们选择基于DEA方法的区域输入效度评价模型. . 它是对多个输入和多个输入的相对有效性或有效性的系统分析在同一类型的单位。首先考虑将该地区作为同类型的决策单位，转换资源消耗（国内生产总值、高等教育学生比例、就业率、互联网用户的比例）转化为产出（第一语言增长率和第二语言增长率）。通过该区域的输入和输出观测数据来确定输入是否有效，从而评估该地区的效率。该模型对投入无效的领域具有诊断效用，提供了改进的证据。



一锤定音：论文印象

论文详细点评请见视频

《美国数学建模竞赛》

完整课程请长按下方二维码

