摘要

我们的目标是调查全球语言的发展趋势和为大型跨国公司建立开设国际办事处的标准，为其挑选6个甚至是更少的国际办事处，使员工能用英语和另外一种或多种的语言进行交流。

1.模型一仅考虑地理因素，模型将选定地区划分为多个语言区，语言区内国家以同种语言作为母语，选定语言区内该语言的发源国家首都为热点城市，将热点城市相连，抽象为一个图模型。（图中连线的标准为，两语言区陆地毗邻才可相连。）我们用距离衡量定义的传播因子，以传播因子作为权重衡量一种语言在不同地区的分布。

2.由于仅考虑地理因素不够全面，在模型二中，我们考虑经贸关系，将全球经济实力排名前九的国家作为九大经济体，每个经济体作为独立的语言区有自己的母语，通过模型预测出经济体间的影响因子，得到一语言区包含的使用不同语言作为第二语言的比例，进而得到一个地区第二语言的分布情况。

3.我们通过灰色预测模型预测出未来50年前20大语言的母语人口总数，根据不同情况以模型一和模型二为基础，计算出使用前20大语言中作为第二语言的人口总数，通过与前20大语言的母语人口总数相加，预测总语言说话人口的总数。

4.建立人口迁移模型，经过分析得到语言地理分布的变化。也就是说，我们通过灰色预测模型预测未来50年人口的净迁入量，迁入量为正 说明此语言区是移民的目的地；如果为负，说明此语言区是移民的源地。从而可以分析出人口的迁移模式，得出不同语言地理分布的变化。

5.为了开设6个国际办事处，需要评估语言的影响力（评估等级），我们制定了指标包括地理，经济，交流，知识与媒体，外交。而且，每个指标包括几个次要指标。由于此问题具有主观性，我们采用层次分析法确定每项指标的权重，得出一个三层分层结构。通过该模型计算不同语言的国际影响力，作为开设办事处的依据。

6.随着通讯水平的发展，距离已经不再是语言交流的鸿沟。我们考虑评估不同国家的通讯水平确定该国家是否需要建立国际办事处，通讯水平越高，沟通越方便，建立办事处的权重较小，通讯水平越低，沟通不便，建立国际办事处的需求较大，建立办事处的权重较大。

引言

一家跨国公司为了拓展国际业务，需要开设用不同语言交流的国际办事处，以理解不同顾客的需求，促进公司发展。

我们主要面临5个问题：

1.模拟随时间的推移各种语言说话者的分布。

2.预测未来50年母语人士、说话人总数发生的变化以及现十大语言中的在未来是否会被另一语言取代。

3.预测全球人口和迁移模式，判断语言的地理分布。

4.为该跨国公司选择六个国际办事处。

5.根据全球通讯性质判断开设国际办事处的数量是否可以减少。

为此，我们建立了地理和经贸关系模型解决语言说话者分布问题。通过灰色预测模型预测出

未来50年母语人士、说话人总数并与现状对比。使用人口迁移模型判断语言的地理分布。运用层次分析法确定5大指标作为选择6大国际办事处的依据。评判各国通讯水平判断开设的国际办事处的数量是否可以减少。

模型一

介绍

我们主要考虑分别建立基于地理分布和经贸关系的两个模型，以解决随着时间的推移模拟各种语言的说话人分布问题。

假设

假设国家之间的距离是根据资本的距离。

假设2017-2025年各国首都不会改变。

假设2017-2025年期间所有国家都不会被诸如战争等不可抗拒的因素所剥夺。

定义

表一

|  |  |
| --- | --- |
| 表示 | 定义 |
| R | 地球半径 |
|  | I语言区经度 |
|  | I语言区纬度 |
|  | i语言区与j语言区关于地球球心的夹角 |
|  | I、j语言区间的距离 |
|  | 两个语言区间的最大距离 |
|  | 第二语言的传播因子（i、j语言区的传播因子相同，因为在一条连线上，括号内便于理解，不翻译） |
|  | J语言区使用i语言作为第二语言的人口数 |
|  | I语言区的人数 |

各个语言区一般都有自己的母语，该图模型上直接标注作为各语言区的母语分布。该模型以地理分布为重点，定义相邻且距离最小为最优的原则，根据计算各个语言区间的距离，考察各个语言区是否相邻，算出在X区域内第二语言的传播因子，并与这个语言区的人数做乘积，得到其他语言区学习本语言的第二语言人数，判断第二语言的分布情况。

模型从各国首都之间的距离开始。 通过Google Earth，我们可以准确定位经纬度。 根据两点的经纬度和距离换算公式（1）得到不同国家之间的距离D\_ij：

=sin()\*sin()\*cos(-)+cos()\*cos()  
= (1)

通过最短路径算法找出不同国家的首都之间的最小距离。 如果你需要计算你想把j作为第二语言学习的国家的数量，那么你需要定义一个第二语言传播因子（2）作为权重。

= (2)

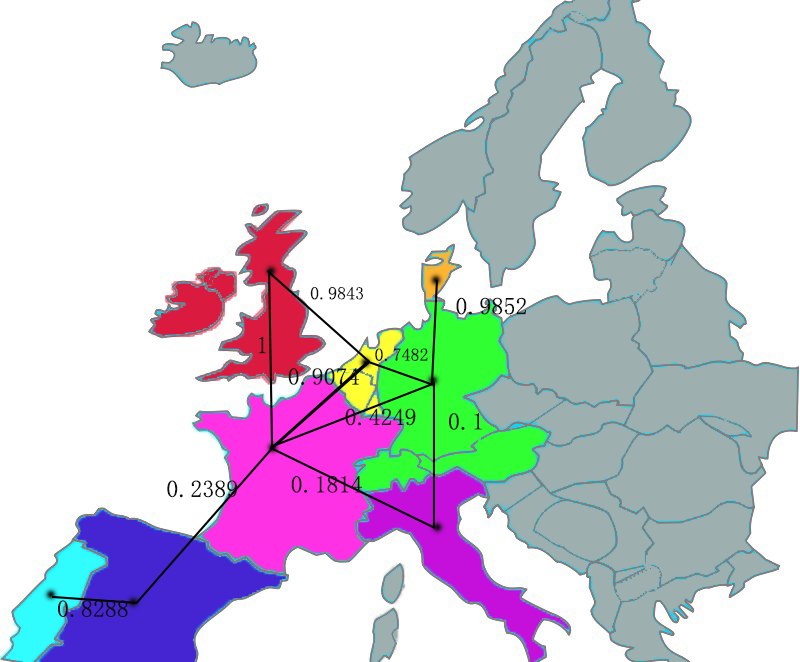
模型将不同时期的语言区域i的人口作为学习语言j的基数与传播因子（2）相乘，并将其用作不同时期的第二语言人口。

= (3)

最终，得到其他语言区学习j语言的第二语言学习者的人数，判断j语言作为第二语言的分布情况。

模型测试

我们选取西欧、除希腊以外的南欧国家及丹麦（因为与德国陆上接壤，所以也选取此国家），由于英国、法国、葡萄牙、西班牙、丹麦、意大利都有自己单独的母语，因此将这六个国家的首都分别作为图模型上的六个点。由于荷兰、比利时的母语都为荷兰语，德国、瑞士、奥地利母语都为德语，所以将荷兰、比利时化为一个荷兰语区，将德国、瑞士、奥地利化为一个德语区，并将荷兰、德国首都作为图模型上的两个点。也就是说，该图模型中以八个语言区为基础进行计算。该图模型如下图所示。



QW`4C_ES)NS7KQGC)`0)EYCPortugal

GL7IY)FHBCIQNVX2$3$7]PJSpain

R[)Y${7I%D)]PYSUBQ]}05JItaly

)KUUS7]RJGM6XXJWP0O]8[5France

)_I$DTC7R17F)6M9Q8)$@]QDutch area

I0]DQTW7S_FXT06X5F9V9LFUnited Kingdom

DZ[{5D7_YC9$]S}R)COO5IFDenmark

C:\Users\lenovo\AppData\Roaming\Tencent\Users\277172705\QQ\WinTemp\RichOle\ANZ82W`NYSR)~J$XALQ@9)4.pngGerman area

首先，我们通过（1）（2）公式计算得到如图所示的第二语言的传播因子，作为权重。

其次，我们在网站上找到了8个语言区在2010年-2050年的人数，通过公式（2）（3）计算分别得到他们作为第二语言的总使用人数。（如下表所示）

(此处插表)

由此，我们可以得到i语言区在其他语言区的分布情况，我们以英语为例

（此处插图）

由此，我们能够看出英语的分布情况，其他语言的分布情况可由此分析过程一一得出。

|  |  |
| --- | --- |
| Symbol | Definition |
| Constant | |
|  | m语言区进口总量中n语言区所占份额 |
| Variables | |
|  | M语言区的人口数 |
|  | M语言区使用n语言作为第二语言的人口数 |

经济发展是国家强大的支柱，而经贸关系是衡量一个国家或地区对于他国经济影响力的重要指标，我们的模型II以经贸关系为重点，综合不同网站上的排名，通过将某个经济体对其他经济体的进口份额以及该经济体地区的人口为衡量指标，建立我们的最强经济影响模型，通过该模型衡量第二语言学习者的总人数。

确定n经济体的人口数，对其取对数并定义其为，所得到的数据易消除异方差问题；同时，取对数以后，变量具有弹性。通过与m经济体中占n经济体的外贸出口份额作乘积，得到

=\* (8)

最终，得到其他语言区学习j语言的第二语言学习者的人数，判断j语言作为第二语言的分布情况。

模型测试

模型三

解决的问题

预测未来50年母语人士和说话人总数情况。

灰色模型预测母语人口数据

我们已经得到过去50年的前20大语言的母语人口数据（每5年统计一次）。由于数据的统计年限是间隔的，每5年统计一次，移民数据间无较强的联系，不可使用线性规划模型。而神经网络模型需考虑过多的因素，模型较为复杂，我们也不考虑。鉴于目前的情况，我们设计一个灰色预测模型获得更高可靠性的数据。使用灰色预测模型的优点是在缺乏可访问的数据的前提下可以得到更可靠的结果，完全符合我们目前的情况。

Indruction

根据过去五十年前20大语言的母语人口数据，采用灰色预测模型预测出预测出未来五十年20大语言的母语人口数。由于总人数的缺少过去年份的数据不能直接预测，考虑用模型一、二分情况分别计算出前20大语言作为第二语言的人口数，并将其加上其各个语言国家的母语人数，作为预测的总人口数。

相关性分析

未来50年母语人口数预测基于Verhulst模型

在灰色预测中，我们试图找到和掌握过去五十年前20大语言的母语人口数据的发展规律，最后通过原始数据对系统的未来状况进行科学的定量预测处理和灰色模型建设。 目前，灰色预测模型GM（1,1）是主要的应用灰色预测，但GM（1,1）模型适用于序列强指数级，只能描述单调的变化过程。人口数量是具有一定随机波动性的动态时变系统，因此更适合于我们使用Verhulst模型进行非单调摆动发展序列。

模型解决方案

X(0)的定义有变动

Results and Conclusions

地理模型是一个国家对多个国家的影响，而经贸关系模型是多个国家对一个国家的影响。在前20大语言中，由于英语、法语、西班牙语的第二语言分布具有分散性，且影响范围较广，主要是其他国家学习它们作为第二语言，是一个母语国家对其他多个国家的影响，所以采用地理模型得到学习英语、法语、西班牙语的第二语言人数。其余语言基本只有本国使用，对他国的影响较小，主要是其他国家对于本国的影响，所以采用经贸关系模型。

根据灰色预测得到的数据，如下图所示

我们可看出2050年母语人口和总说话人口的变化

模型四

数据

1. 我们找到了联合国提供的未来50年全球人口数，

（插入表）

2. 世界银行提供了过去五十年各国人口净迁入量，用灰色预测模型预测未来五十年人口净迁入量。

（插入表）

Assumption

假设迁入量为正说明此语言区是移民的目的地；假设为负，说明此语言区是移民的源地，从而可以分析出人口的迁移模式。

由此，可以分析出人口的迁移模式。

|  |  |
| --- | --- |
| 定义 |  |
|  | i语言作为母语的一个国家的净迁移量 |
|  | 是i语言作为母语的所有国家净迁移量的累加 |
|  | 总拉力因子即移民的方向 |
|  | 地缘拉力因子 |
|  | 经济拉力因子 |
|  | 影响地缘拉力因子的权重 |
|  | 影响经济拉力因子的权重 |
|  |  |

是i语言作为母语的一个国家净迁移量。迁出人口本身不改变母语，迁入国语言是其第二语言，但会对迁入国家的语言造成影响。是用来判断i语言说话人口的迁移模式（源地或目的地），

= m1+ m2 + …..+ ，其中n是某语言作为母语的所有国家数

总拉力因子代表移民的方向，从地缘和经济的角度考虑。包括地缘拉力因子和经济拉力因子。其中，以权重和判断两个因子在中的所占比重。

语言区i是移民的源地，语言区j是移民的目的地，表示语言区i到语言区j的迁移人口

=

统计出迁出语言区，迁入语言区，只考虑迁入语言区的来向，计算得到top14

模型五

概述

在一个多语言的全球化社会中，语言是让我们能够与他人交流、允许我们参与社会的文化、经济等活动的首要前提。那么哪些语言是最有用的？现在有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，在全球不同国家分别建设6个办事处，你会在这6个办事处中使用哪种语言？这将是本模型中我们要解决的问题。

该模型采用层次分析法衡量语言的“有用性”，我们定义了一个语言能力指数，分配权重。广义来说，主要有5种因素影响语言的有用性。分别为：

1.地理

2.经济

3.交流

4.知识与媒体

5.外交

从狭义上说，5种因素分别由不同指标衡量（参见下表）。也就是说，语言能力指数（PLI）采用17个指标评价语言的影响力，即权重。该指数评价的是语言对人类整体而言的有用性，而并非针对受到任何地理环境、人文情况和个人偏好影响的单一个体。这一指数也无法衡量语言和其相关的文化的优美程度和价值。

（此处插入17个指标的表）

进而对我们的模型进行敏感性分析得到，分析不同的结果并为理想的结果找到更好的参数。综上所述，我们的模型是一个具有技术和数据支持的可行和合理的模型。由于其主观性，数据训练后可以灵活运用。

选取十七大指标的原因

语言能力指数（PLI）是衡量语言影响力的标准，在这里，我们用五个因素来衡量PLI：地理，经济，交流，知识与媒体，外交。但为什么使用这五个因素？

首先，五个因素之间的相关性很低，每个因素都可作为PLI的影响指标分离出来。

其次，它们与PLI密切相关：

（a）地理：地理空间距离发挥着它的空间隔离作用，相同国家或相邻国家说同一种语言的可能性更大。有三个指标代表地理，分别为地理面积，临近国家数和语言使用国数。

（b）经济：经济体通过自身经济实力产生的影响力，影响着其他语言区域国家语言政策的设定和语言培养体系的建立，同时，经济影响着人们对语言学习的选择，从而推动某种语言的传播速度和范围。有四个指标代表经济，分别为国家GDP,人均GDP，出口，外汇。

（c）交流：语言的交流促进语言的发展，母语使用人数与第二语言使用人数是衡量语言影响力的重要指标，同时，随着旅游业的迅速发展，促进了语言的交流与学习，也扩大了某种语言的影响力。也就是说，母语使用人数，第二语言使用人数和出境旅游的人次是衡量交流因素的指标。

（d）知识与媒体：随着互联网的迅速发展，网络使地球称为了一个地球村，互联网用户与日俱增，增大了不同国家人民沟通交流的可能性，进而扩大了不同国家语言的影响力。同时，教育水平是衡量人民对于学习不同种类语言可能性的因素，该模型用全球top200大学数与教育学投资数作为指标衡量教育水平进而判断某种语言的“有用性”。

（e）外交：一个国家的外交程度影响着该国第二语言的学习，其中，我们以加入国际货币基金组织的国家，联合国常任理事国国家，加入世界银行的国家和十个国际联盟为参考指标判断各个语言的“有用性”。

解决步骤：

当我们试图获得一级的五个方面的权重评估和十七个二级评估标准的权重，主观判断是不恰当的。 所以我们选择Analytic Hierarchy过程（AHP）作为组合所有权重系数的方式评估系统中的指标。

确定判断矩阵。我们采用成对比较法和层次分析法1-9构造判断矩阵

计算特征值和特征向量。矩阵A的最大特征值λ\_max具有相应的特征向量u

做一致性检查。一致性的指标是

其中n是矩阵的维数。一致性比率的表达是

确定了评价体系中所有指标的权重系数后，我们量化了…的重要性.

CWi表示标准水平因子i的权重，其中AWj是第i个临界水平的次级临界水平因子j的权重，mi表示次级临界因子的总数，Fj表示次级临界水平因子。

最后，我们可以得到...使用AHP模型。

由AHP计算得到前十大语言国家的语言能力指数排名，如下表所示

结论

由上表分析可得到如下结论：由于纽约和上海已经建立有该公司的国际办事处，因此，不再在这两个地区语言使用国建立国际办事处。我们只考虑上表中排名为3-8的语言区建立国际办事处，建立的具体位置为该语言区的中心热点城市，分别为巴黎，马德里，吉达，莫斯科，慕尼黑，东京。

模型六

考虑到每个中心的热点国家的通信水平，并与选定的办公地点进行比较，以确定是否需要减少办公室的数量。

|  |  |
| --- | --- |
| Symbol | Definition |
|  | Internet hosts |
|  | Internet users |
|  | Internet level |
|  | Telephone users |
|  | Maximum telephone users |
|  | Mobile communication level |
|  | The weight of Internet level |
|  | The weight of mobile communication level |
|  | Communication level |

我们采用两种评价方式，第一种用来衡量互联网水平：

=

第二种用来衡量移动通信水平

M=

=

比较办公室所在国家i的通信水平，用如下公式衡量=0.8,=0.2

+

通过上述公式，我们仅在考虑通信水平的情况下得到前十大语言的排名，如下表所示：

上表是将模型5中选出的六个国际办事处所在国的通讯水平排列，水平越高，网络通讯越发达，从节约公司成本的角度出发，可将其删去。

结论

为了解决该跨国公司的问题，团队需要描述调查全球语言的发展趋势以及国际办事处的选择的模型，我们对于各个分问题逐一解决，

对于解决模拟随时间推移各种语言说话者的分布问题，团队综合地理、经贸关系两因素，建立了衡量一种语言在不同地区的分布情况与一个地区第二语言的分布情况的模型。为解决预测未来50年母语人士和说话人总数问题，团队使用灰色预测模型得到结果。为解决语言地理分布的变化问题，团队建立人口迁移模型，得出不同语言地理分布的变化。跨国公司需要我们设计模型建立6个国际办事处，为此我们选择AHP层次分析法进行建模，得出前十大语言的语言能力指数作为判断语言影响力的因子，选取办事处，并以国家通讯发展水平为依据，为节约公司成本，尽可能减少国际办事处的数量。