目录

[1 一级标题 2](#_Toc11615486)

[1.1 二级标题 2](#_Toc11615487)

[1.2 二级标题 2](#_Toc11615488)

[2 EMD算法概述 2](#_Toc11615489)

[2.1 EMD的概念和基本算法 2](#_Toc11615490)

[2.1.1 EMD算法的基本概述 2](#_Toc11615491)

[2.1.2 EMD算法的停止准则 3](#_Toc11615492)

[2.1.3 EMD算法的端点问题 4](#_Toc11615493)

[2.2 二级标题 4](#_Toc11615494)

[3 第三章 4](#_Toc11615495)

[3.1 二级标题 4](#_Toc11615496)

[3.2 二级标题 5](#_Toc11615497)

[3.2.1 三级标题 5](#_Toc11615498)

1 一级标题

1.1 二级标题

1.2 二级标题

2 EMD算法概述

经验模式分解(Empirical Mode Decomposition, EMD)是由Huang(1998) [73]等于1998年提出的一种处理非线性、非平稳性信号的方法, 其本质是对时间序列进行平稳化处理。与小波分解不同的是EMD是由数据驱动的一种序列分解方式，其更具原序列的极值特征直接对原序列进行分解不依赖于基函数。EMD本质上可以被看作是一种滤波，将不同尺度的波动逐级进行分解。经验模式分解因为其前提假设少，适用性广，提出之后很快就被用于许多理工科领域[59]。

2.1 EMD的概念和基本算法

2.1.1 EMD算法的基本概述

首先Huang在其文章中定义本征模态函数(Intrinsic mode functions)。当函数满足以下两个条件时，我们视其为一个本征模态函数：

（1）在整个数据集中，极值点的个数和过零点的个数必须相等或最多相差1个；

（2）在任意时间点，由局部极大值定义的包络线和由局部极小值定义的包络线的均值为零。

EMD分解是一种直观的、后验的、自适应的分解，完全有数据驱动。该方法的实质是对数据中固有振荡模态的特征时间尺度进行经验识别，然后对数据进行相应的分解。EMD分解基于以下假设：

（1）信号序列至少有两个极值，一个极大值一个极小值；

（2）时间尺度的特征由极值之间的时间间隔来定义；

（3）如果数据没有极值只有拐点，那么可以通过对数据差分来获取极值点。

从本质上讲经验模式分解过程是一个筛选过程，从该过程中依次筛选出频率由高到低的本征模态函数，并最终剩余一个不能再分解的单调残差序列，也称为趋势项。也就是说EMD可以自适应地把一个时间序列分解成有限个IMFs和一个残差项。算法描述如下：

Step1：输入原始序列信号；

Step2：获取原始序列的所有极大值点和极小值点，并分别对极大值点和极小值点用2条3样条插值曲线拟合，得到一条上包络线（upper envelope）和下包络线（lower envelope）。

Step3：计算包络线的均值；

Step4：计算和的差别并令；

Step5：检查是否满足上述的两个IMF的条件，如果是一个IMF或满足停止准则，则转到Step6；否则令，并重复Step1-4直到是一个IMF或满足停止准则。

Step6：计算残差项（residue），如果是一个单调函数或者只有一个极值点则分解完成；否则令，重复Step1-5。

至此，我们将原始信号分解为若干个本征模态函数和一个趋势项，用公式（2-1）表示：

 公式（2-1）

在上述的分解过程中，分解得到的本征模函数很容易满足第一个条件，即极值点的个数与过零点的个数相等或最多相差1个，然而第二个条件即局部上包络线和下包络线的均值为0却不一定能通过变换满足。而且当第二个条件发挥到极致时，可能会消除物理意义上的振幅波动，因此在筛选过程中要格外小心，因为如果将筛选进行到极致，可能使得到的IMF成为恒定的纯调频信号。为了确保IMF各组成部分保持足够的振幅和频率调制的物理意义，我们必须确定一个停止筛选过程的标准[73]。

2.1.2 EMD算法的停止准则

根据2.1.1中IMF的定义，严格IMF应满足局部极大值定义的包络函数和局部极小值定义的包络函数的均值恒为0，而在实际分解过程中很难做到分解出的上下包络之和严格等于0，只能接近于0。针对这种情况，一些学者通过实验给出了一些相对较为理想的停止准则也被称为局部包络条件。

EMD的提出者Huang(1998)[73]在其论文中给出了标准偏差(SD)准则，通过限制SD的值来判断是否停止筛选。SD是由两个连续的筛选结果计算得到的，计算过程用公式（2-2）表示[73]：

 公式（2-2）

其中为第个IMF第次迭代的值，计算公式如见2..1.1。当SD小于研究者给定的某个值时，就认为该为第个本征模态函数。Huang在其文章中通过大量实验给出取值在0.2和03之间时分解效果较好。后来Huang等人又给出了另一个相似的停止准则用公式（2-3）表示：

 公式（2-3）

其中为第个IMF第次迭代的上下包络线的均值（计算见2.1.1）。上述两种停止准则由柯西收敛性检验给出在数学上似乎严格。然而，由于以下原因很难实现这个标准:首先，SD小到什么程度就足够小，这其实并没什么标准。其次，这个准则不依赖于IMFs的定义，因为平方差很小但不能保证函数会有相同数量的过零点和极值。为了弥补这些缺点，Huang(1999,2003)等人提出了第二种类型的标准，称为S停止准则。使用这种类型的停止准则，筛选过程只有在零点个数和极值个数相等或最多相差1且连续S次保持相同时才停止。Huang(2003)等人的大量试验表明，S的最佳范围应该在3到8之间，但更倾向于较低的数字。显然，任何选择都是临时的，需要一个严格的理由[74]。

法国学者Gabriel Rilling(2003)[75]等基于Huang等的研究提出了一个较为常用评价函数来控制筛分过程的停止

 公式（2-4）

其中为上下包络线的均值，。并设置三个阈值、、，当满足以下两个条件：

1. 当里面小于的比率达到；
2. 不存在大于的值。

则认为分解结束，中止筛选，一般取，和[75]。这也是本文中所采用的停止准则。

2.1.3 EMD算法的端点问题

在EMD相关算法的应用过程中，端点效应的处理是数据预处理过程中一个重要的步骤。所谓端点效应，是指在运用EMD对信号进行分解时，需要去“筛选”出本征模函数，而“筛选”过程需要对极值点使用三样条插值构造上下包络线，如果数据在端点处不是极值，那么在对极值点插值拟合时，将会出现误差从而产生端点效应。在端点附近时，三样条插值将出现发散现象，而随着本征模函数的一步一步提取，端点效应向内扩散最终导致污染数据[59]。Huang在其文章中指出在信号的两端，根据端点信号的振幅和频率，分别加两个特征波，可以有效抑制端点效应，但没有给出具体的做法。

处理端点问题最简单的方法是在迭代过程中不断地删除非极值的部分，但是对于预测问题来讲抛弃右端点的值也就意味着抛弃了近期的数据，然而近期的数据包往往是含着信息最大并且最重要的数据。所以如果是对时间序列值做简单的分析分解的话可以使用此方法，但是对于预测问题，往往不适合此方法。本文主要研究的是对股价指数的预测问题，不适合使用此方法。

另外一种解决端点问题的思想就是人为的添加一些数据，也称为延拓法。随着EMD算法的广泛应用，一些学者陆续的提出了一些较为经典的延拓方法，这些延拓方法大致可以概括为两种直接延拓法和预测延拓法[59]。

直接延拓法就是通过一定的技术手段将原有数据向外延拓。目前主要的延拓方法有，镜像延拓[76]、偶延拓、周期延拓[80]、极值延拓、平行线延拓、加窗口延拓[81]、、Coughlin法（Coughlin and Tung 2004）[79]、斜率法(Datig and Schlurmaim 2004)[77]和Rato法(Rato 2008)[78]等。

预测延拓法主要是指根据各种预测模型或方法对未来一些点进行外推预测的方法。常见的预测延拓方法有时间序列预测延拓，神经网络预测延拓，以及一些利用机器学习方法进行端点预测的方法。与直接延拓相比预测延拓往往需要更大的计算量，而且要去根据具体情况选择合适的预测方法，这样给延拓带来个一些不确定性。

本文主要研究的是模型的简化和优化问题，并不关注这些延拓方法的区别。所以本文取最为常用并且简单的延拓方法-镜像延拓法来处理端点问题。镜像延拓法通过对原始数据序列邻近边缘的极值点添加镜像,以对称的方式增加n个极值点的方式来抑制分解过程中的端点效应。对于图2.1镜像延拓法过程如下：

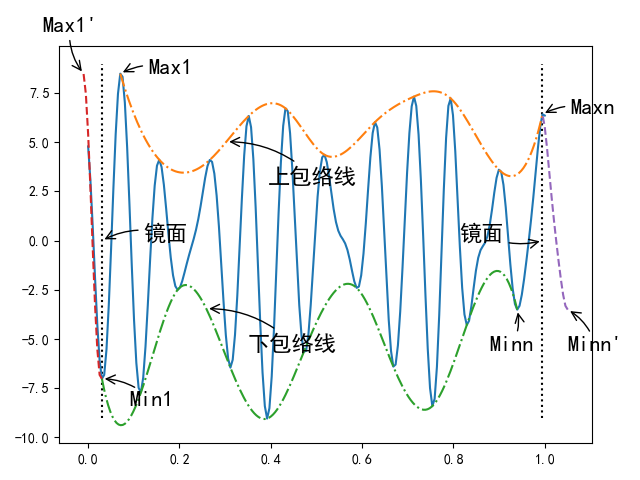


图2.1 镜像延拓法过程

Step1：找到原始序列最左端的极值点Min1（第一个极小值点），以及最右端的极值点Maxn（最后一个极大值），如图2.1中所示；

Step2：别以Min1和Maxn为镜面得到Max1的镜面对称点Max1’和Minn的镜面对称点Minn’,如图2.1所示；

Step3：最后根据数据原有的极大值点和极小值点加上有镜面对称得到的点，分别对运用三样条插值构建上包络线和下包络线。这样可以在一定程度上一直端点的发散（图2.1给的上、下包络线为未进行延拓下的包络线）。

2.2 EMD算法的改进研究

因为停止条件的限制，在经验模态分解过程中可能会出现模态混叠，使得不同特征的序列分解成一个子序列。针对此类问题一些学者提出了一系列的改进算法。

2.2.1 集成经验模态分解

针对经验模态分解存在的模态混叠现象，Wu and Huang(2009)[82] 等人提出集成经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)算法，显著地提高了经验模态分解算法的应用价值。

集成经验模态分解（EEMD）算法的主要思想是：向原始的序列数据添加一个白噪声，然后对白噪声和原始序列的混合序列进行分解，重复N次此过程，将所得到的N组IMF和N个余项分别取平均值作为最后的分解的结果。集成经验模态分解过程中认为经过N次集成平均，序列中添加的白噪声可以相互抵消，同时克服了经验模态分解中出现的模态混叠现象，能够更好的实现信号序列的分解，是经验模态分解算法发展中的一个极其重要的改进。

在2.1.1章节EMD算法的基础上，EEMD算法步骤如下：

Step1：在原始信号中添加一个白噪声信号，得到下列混合信号；



其中，表示第次信号混合，，为信号混合的总次数，是一个人为设定参数，通常取[82]。

Step2：对混合后的信号序列进行经验模态分解，得到第次分解的子序列和其中为分解的个IMF（为了确保分解信号个数相同一般要求添加的白噪声为低信噪比的白噪声序列）。重复该步骤，直至。

Step3：对每个IMF和余项取平均，作为最终分解结果，得到第项IMF和余项：





Step4：得到最终的分解结果：



其中，为分解的到的本征模函数的个数。

2.2.2 完整集成经验模态分解

根据2.2.1中的介绍，我们知道集成经验模态分解算法是通过添加噪声项来解决模态混叠问题。在引入噪声项后我们假设经过多次添加噪声项，噪声之间可以相互抵消。但实际上在信号重构时，引入的噪声项并不能完全抵消，一般都是通过增加集成的次数来减弱残留噪声项的相对大小，，但是这样做往往会大大的增加计算量。为解决经验模态分解过程中的模态混叠问题，同时消除集成分解过程中添加的白噪声对信号重构的影响, Yeh J R and Shieh J S(2010)[83]提出了完整集成经验模态分解(Complete Ensemble Empirical Mode Decomposition, CEEMD)。该算法的主要思想是成对的添加噪声，并且使得所添加的白噪声和为零。基于EMD和EEMD算法，CEEMD算法分解过程如下。

Step1：将白噪声成对的添加到原始信号中，得到一组待分解信号，并且该组信号的白噪声序列和为零：



其中，表示原始信号序列，表示加入的白噪声，表示与正的白噪声叠加后的信号序列，表示与负的白噪声叠加后的信号序列。

Step2：分别对、进行EMD分解，重复N次。得到分解后的IMF子序列和余项：





其中，、和、分别表示第次混合信号、分解得到的子序列和余项。

Step3：最终，由CEEMD算法将原始信号分解为



其中，为分解的到的本征模函数的个数。

2.2.3 具有自适应噪声的完整集成经验模态分解

EEMD和CEEMD都是通过对原序列添加噪声来处理模态混叠问题。有些学者考虑在分解过程中添加噪声来处理模态混叠问题，进而提出了具有自适应噪声的完整集成经验模态分解(Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, CEEMDAN)算法，该算法的主要思想是：在序列分解过程中的每个阶段都添加自适应白噪声，通过计算唯一的余量信号获得IMF子序列，在较少的集成次数后就能够使重构误差几乎为零[84]。CEEMDAN算法过程如下（详见Humeau-Heurtier(2105)[85]）：

Step1：在原始信号序列中添加一些自适应白噪声：



其中，表示第次加入白噪声后的信号，表示噪声系数（噪声标准差），表示第次添加的白噪声，表示集成次数。

Step2: 对每个信号，分别采用EMD算法分解获得第一个IMF分量，然后对这个分量取均值：



其中，表示分解得到的第一个IMF信号。则余项信号为：



Step3：定义为使用EMD算法分离出第个本征模函数的过程，分解，，直到分解出一个IMF。(在该步)在每一步允许选择的信噪比（SNR）。然后，得到第二个IMF：



Step4: 计算第个余项：



使用EMD算法分解，直到分解出一个IMF，定义第IMF：



Step5：回到Step4，并令，直到余项不能再被分解。得到最终的余项：



其中为分解得到的IMF个数。最终原信号序列被写作：



2.3 本章小结

本章对经验模态分解算法进行了详细的介绍。介绍了经验模态分解算法的思想、基本原理以及算法实现过程。而且本章也对经验模态分解存在的问题以及其改进算法做了一些简单的介绍。本文并没有深入探讨不同的端点、不同的经验模态分解算法

3 第三章

3.1 二级标题

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model |  | Lag | | | | | | | |
|  | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean |
| ARIMA | RMSE | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 |
|  | MAE |  |  |  |  |  |  |  |  |
|  | RMSE |  |  |  |  |  |  |  |  |
| ANN | MAPE | 1.381 | 1.377 | 1.378 | 1.377 | 1.378 | 1.381 | 1.378 | 1.379 |
|  | MAE | 39.58 | 39.45 | 39.49 | 39.44 | 39.49 | 39.57 | 39.47 | 39.50 |
|  | RMSE | 52.82 | 52.65 | 52.70 | 52.62 | 52.68 | 52.81 | 52.65 | 52.70 |
|  | （1,T） | 0.071 | 0.072 | 0.074 | 0.076 | 0.071 | 0.074 | 0.074 | 0.073 |
| E-ANN | MAPE | 0.687 | 0.673 | 0.690 | 0.688 | 0.695 | 0.696 | 0.696 | 0.689 |
|  | MAE | 19.77 | 19.39 | 19.88 | 19.83 | 20.06 | 20.09 | 20.07 | 19.87 |
|  | RMSE | 26.04 | 25.95 | 26.45 | 26.30 | 26.44 | 26.53 | 26.42 | 26.30 |
|  | （13,T） | 0.943 | 0.932 | 0.963 | 1.002 | 0.968 | 0.956 | 0.954 | 0.96 |
| S-E-ANN | MAPE | 0.673 | 0.684 | 0.683 | 0.678 | 0.676 | 0.671 | 0.673 | 0.677 |
|  | MAE | 19.39 | 19.72 | 19.68 | 19.53 | 19.50 | 19.40 | 19.46 | 19.53 |
|  | RMSE | 25.29 | 25.59 | 25.52 | 25.91 | 26.06 | 26.01 | 26.27 | 25.81 |
|  | （1,T） | 0.082 | 0.085 | 0.089 | 0.094 | 0.095 | 0.096 | 0.097 | 0.091 |

E-A表示EMD-ANN，S-E-A表示S-EMD-ANN，E-L表示EMD-LSTM，S-E-L表示S-EMD-LSTM，（N，T）表示模型个数和训练模型所需时间

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model |  | Lag | | | | | | | |
|  | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean |
| ARIMA | RMSE | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 | 34.45 |
|  | MAE |  |  |  |  |  |  |  |  |
|  | RMSE |  |  |  |  |  |  |  |  |
| ANN | MAPE | 1.039 | 1.036 | 1.039 | 1.038 | 1.035 | 1.038 | 1.037 | 1.038 |
|  | MAE | 28.06 | 27.98 | 28.05 | 28.02 | 27.96 | 28.04 | 28.02 | 28.02 |
|  | RMSE | 39.86 | 39.75 | 39.81 | 39.79 | 39.68 | 39.78 | 39.76 | 39.78 |
|  | （1,T） | 0.072 | 0.073 | 0.072 | 0.072 | 0.072 | 0.072 | 0.072 | 0.072 |
| E-ANN | MAPE | 0.519 | 0. 528 | 0.529 | 0.525 | 0.527 | 0.528 | 0.530 | 0.527 |
|  | MAE | 14.02 | 14.27 | 14.28 | 14.19 | 14.24 | 14.27 | 14.32 | 14.23 |
|  | RMSE | 19.38 | 19.88 | 19.81 | 19.79 | 19.81 | 19.88 | 19.89 | 19.78 |
|  | （12,T） | 0.868 | 0.869 | 0.873 | 0.868 | 0.869 | 0.870 | 0.876 | 0.879 |
| S-E-ANN | MAPE | 0.511 | 0.514 | 0.516 | 0.521 | 0.523 | 0.518 | 0.524 | 0.518 |
|  | MAE | 13.84 | 13.93 | 13.96 | 14.11 | 14.16 | 14.04 | 14.19 | 14.03 |
|  | RMSE | 19.21 | 19.10 | 19.12 | 19.28 | 19.45 | 19.40 | 19.64 | 19.31 |
|  | （1,T） | 0.083 | 0.086 | 0.087 | 0.089 | 0.091 | 0.094 | 0.095 | 0.089 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model |  | Lag | | | | | | | |
|  | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean |
|  | DM | 0.377 | 0.904 | 0.365 | 0.409 | 0.212 | 0.220 | 0.238 | 0.413 |
|  | WX | 0.554 | 0.801 | 0.742 | 0.615 | 0.328 | 0.483 | 0.436 | 0.651 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model |  | Lag | | | | | | | |
|  | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean |
|  | DM | 0.435 | 0.211 | 0.203 | 0.316 | 0.243 | 0.225 | 0.185 | 0.272 |
|  | WX | 0.943 | 0.640 | 0.890 | 0.932 | 0.809 | 0.828 | 0.712 | 0.882 |

3.2 二级标题

3.2.1 三级标题