基于熵权法-多维云模型的围岩稳定性分类

摘要：基于围岩综合稳定性分类的模糊性，现对已有围岩样本数据，利用熵权法计算得到围岩稳定性评估中的各指标的权重，并通过对已有的一维云模型进行拓展，得到多维度云模型。将以往一维云模型结果中的各指标线性组合调整为多维云模型中的各指标非线性组合，并通过建立两个采用不同指标的评价模型，对比其他方法的评价结果，探究该评估模型的可行性。

0 引言

当前我国正高速发展，日益突出的交通问题与能源问题正在通过高铁、高速公路和地铁等交通工程与采矿工程来逐步缓解，而围岩稳定性评价是这些建设中多数需涉及的岩石地下工程必须面对的关键问题，合理正确的围岩稳定性分类将影响工程的设计施工及造价。连建发通过可拓方法建立了地下工程围岩评价的物元模型，通过关联函数确定各评价指标权重并计算围岩的稳定性评级，机器学习方法中的随机森林算法，支持向量机和人工神经网络等算法通过刘阳、赵婷和周盛全等人也应用到了围岩稳定性评价中，并得到了比较满意的评价结果，但机器学习算法中需要解决的问题是学习资源的不足，而且这些方法并不能很好地将围岩稳定问题中的模糊性和随机性考虑在内，因此汪明武等人通过建立基于一维云模型的边坡稳定性评价模型，并对围岩样本做出了与其他方法一致的评价结果，其在论文中将云模型对各评价指标的评价结果，通过权重[连建发]进行线性组合，得到各分类等级最终的确定度，并取最大确定度对应的等级作为最终评价等级。而本文方法虽然在核心思想上同一维云模型相似，但通过对一维云模型的拓展，改进云模型参数的选取，将多个评价指标整合到一个模型中，在评价模型的建模工作量上由几十个变为几个，并改用熵权法确定各指标权重，最后以此方法建立新模型实现综合评价结果的计算，并验证该方法的可行性。

1 云模型理论

李德毅等在传统模糊集理论和概率统计的基础上提出了定性定量不确定性转换模型——云模型，并以云模型为基础经过系统研究和发展形成了云理论，其理论能够有效地实现定性概念与定量数据之间的转换，或者简单地说云模型是定性定量间转换的不确定性模型，其在数据挖掘和模糊评价中有着广泛的应用。

1.1云模型定义

设U是一个论域U={x},T是与U相联系的语言值。U中的元素x对于T所表达的定性概念的隶属度CT(x)(或称x与T的相容度)是一个具有稳定倾向的随机数,隶属度在论域上的分布称为隶属云,简称为云，随机数集{x}中的每一个元素x称该云上的为一个云滴。

CT(x)在[0,1]中取值,云是从论域U到区间[0,1]的映射,即:



上述表明云滴集{x}对应一个隶属度集{CT(x)}，而隶属度集的统计学特征即表征了该模型由定量与定性之间的数学关系，云理论中定义了期望Ex，熵En和超熵He这3个特征数据，即表明云滴集与隶属度集关系的统计学特征，也来表征一个云模型。

期望Ex即为整个云滴集{x}在定量上的期望值，该值随着云滴数量的增加无限指向云的中央位置，是最能代表这个定性概念的值；

熵En:是定性概念模糊度的度量,反映了在论域中可被这个概念所接受的数值范围,体现了定性概念亦此亦彼性的裕度。熵越大,概念所接受的数值范围也越大,概念越模糊。

超熵He:是模型的熵En的熵,反映了云滴的离散程度。超熵越大,云滴离散度越大,隶属度的随机性越大,云的“厚度”也越大。

1.2一维云模型

正态云模型是最常用的云滴分布类型，其中每一个云滴都通过云发生器产生，云发生器主要有正向云发生器和逆向云发生器两种。这里只用正向云发生器即可，正向云发生器能够实现定性向定量的转化，即基于云模型的3 个特征数据和云滴数N，计算输出N 个云滴的定量值及由这些云滴构成的云图

其每个云滴生成算法如下：

1、x = G(Ex,En),其中x为云发生器的一次发生结果，且{x}满足x~N(Ex,En)

2、En’=G(En,He), 其中En’为云发生器的一次发生结果，且{ En’}满足x~N(Ex,En)

3、计算x的确定度μ



C:\Users\Administrator\Desktop\figure_3.emf其模型可视化如下图所示：

1.3多维云模型

多维云模型同一维云模型类似，区别在于一维云模型中每个云滴的确定度与一个指标量相关，而多维云模型中每个云滴的确定度受多个指标量影响，即每个云滴的确定度整合了多个指标，这也是在同样评价指标下多维云模型在数量上比一维云模型少许多的原因。

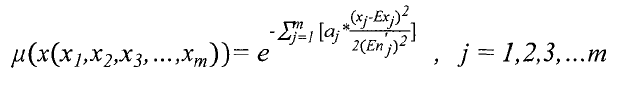
其云模型也是由3个特征数据与云滴数N组成，这3个特征数据不再是单个数值，而是多维向量，其维度与云模型的维度一致。

其每个云滴生成算法如下：

1、对每一维度的特征数据Exi,Eni ,使xi = G(Exi,Eni),其中xi为云发生器的一次发生结果，且{xi}满足xi~N(Exi,Eni)

2、对每一维度的特征数据Eni,Hei,使Eni’=G(Eni,Hei), 其中Eni’为云发生器的一次发生结果，且{ Eni’}满足xi~N(Exi,Eni)

3、计算(x1,x2,x3,…,xm)的确定度μ



上式中{aj}是各评价指标

高维度云图无法可视化，下图以二维云模型为例进行可视化：

C:\Users\Administrator\Desktop\figure_46.emf

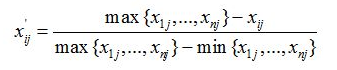
2 熵权法

在信息论中，熵是对不确定性的一种度量。信息量越大，不确定性就越小，熵也就越小；信息量越小，不确定性越大，熵也越大。

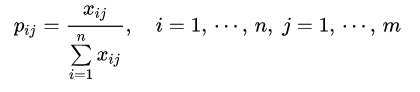
根据熵的特性，可以通过计算熵值来判断一个事件的随机性及无序程度，也可以用熵值来判断某个指标的离散程度，指标的离散程度越大，该指标对综合评价的影响（权重）越大。

熵权法是一种客观赋权法。其计算步骤如下

1. 对 n 个样本， m 个指标，则 xij 为第 i 个样本的第 j 个指标的数值( i=1,2,3,…,m )；
2. 指标的归一化处理



1. 计算第j项指标下第i个样本值占该指标的比重：



1. 计算第j项指标的熵值：

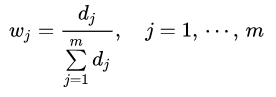




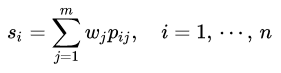
5. 计算信息熵冗余度（差异）：



6. 计算各项指标的权重：



7. 计算各样本的综合得分：



由以上步骤可以得到该样本中各指标权重。

3 围岩稳定性模型的创建

3.1基本过程

通过对每个云模型确定基本参数，一旦参数确定，该云模型即被确定，继而计算实际样本数据对该云模型的确定度，对于围岩稳定等级的划分，可确定五个不同等级的模型，任一样本在每个等级下即可计算出对应的确定度，根据确定度大小即可确定该样本最可能属于哪个等级。其基本计算流程如下：

C:\Users\Administrator\Desktop\云模型分级流程.emf

3.2多维云模型参数确定

围岩稳定性受影响因素很多，依据工程经验和统计分析，得出边坡稳定性主要与地形地貌特征、地质环境特征和气象水文特征等因素有关，这里参照文献[胡宝清]选择评价指标C1 为岩石质量指标(RQD (%)), C2 为湿抗压强度(Rw/MPa), C3 为完整性系数(kv), C4 为结构面强度系数(k f), C5 为地下水渗水量(W/ L·min/10m)，这五项指标作为评价依据，每项指标将指标值按照下表划分为5个稳定等级：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 分类等级 | C1 | C2 | C3 | C4 | C5 |
| 稳定N1 | 90～100 | 120～200 | 0.75～1.00 | 0.80～1.00 | 0～5 |
| 基本稳定N2 | 75～90 | 60～120 | 0.45～0.75 | 0.60～0.80 | 5～10 |
| 稳定性差N3 | 50～75 | 30～60 | 0.30～0.45 | 0.40～0.60 | 10～25 |
| 不稳定N4 | 25～50 | 15～30 | 0.20～0.30 | 0.20～0.40 | 25～125 |
| 极不稳定N5 | 0～25 | 0～15 | 0.00～0.20 | 0.00～0.20 | 125～300 |

云模型的Ex,En及He参数即由上分级表确定，设Smin,ij及Smax,ij为i指标下等级j的指标上下限时，取j等级多维云模型的期望向量Ex,j = {(Smax,ij+Smin,ij)/2}，i=C1,C2,…,C5；j等级多维云模型的熵向量En,j = {(Smax,ij-Smin,ij)/2}，i=C1,C2,…,C5；超熵He体现了某一确定度下指标值的离散程度，这里参考文献[汪明武]取值0.01。

云模型云滴确定度的计算需提供各指标对综合评价所占的权重，本文通过文献[陈守煜]搜集的12个实际围岩样本数据如下表，利用熵权法计算得到权重向量{a}如下：

云模型Ex，En,He参数

|  |  |  |  |
| --- | --- | --- | --- |
| 分类等级 | Ex | En | He |
| N1 | (95,160,0.875,0.9,2.5) | (31.67,53.33,0.29,0.30,0.83) | (0.01,…0.01) |
| N2 | (82.5,90,0.6,0.7,7.5) | (27.50,30.00,0.20,0.23,2.50) | (0.01,…0.01) |
| N3 | (62.5,45,0.375,0.5,17.5) | (20.83,15.0,0.125,0.1667,5.83) | (0.01,…0.01) |
| N4 | (37.5,22.5,0.25,0.3,75) | (12.5,7.5,0.0833,0.10,25.0) | (0.01,…0.01) |
| N5 | (12.5,7.5,0.1,0.1,212.5) | (4.167,2.5,0.0333,0.0333,70.83) | (0.01,…0.01) |

样本实测指标值

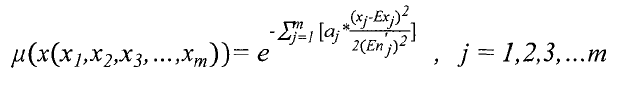
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 样本 | C1 | C2 | C3 | C4 | C5 |
| 1 | 82 | 95 | 0.7 | 0.35 | 20 |
| 2 | 68 | 90 | 0.57 | 0.35 | 20 |
| 3 | 40 | 25 | 0.22 | 0.35 | 20 |
| 4 | 87 | 95 | 0.7 | 0.5 | 10 |
| 5 | 76 | 90 | 0.57 | 0.5 | 10 |
| 6 | 76 | 95 | 0.7 | 0.5 | 10 |
| 7 | 72 | 90 | 0.57 | 0.5 | 10 |
| 8 | 51 | 40 | 0.38 | 0.5 | 10 |
| 9 | 52 | 25 | 0.22 | 0.5 | 10 |
| 10 | 68 | 90 | 0.38 | 0.3 | 20 |
| 11 | 28 | 40 | 0.32 | 0.3 | 20 |
| 12 | 51 | 25 | 0.15 | 0.3 | 20 |

熵权法权重计算结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 量值 | C1 | C2 | C3 | C4 | C5 |
| k | 0.40243 | 0.40243 | 0.40243 | 0.40243 | 0.40243 |
| ej | 0.97296 | 0.928122 | 0.937613 | 0.983239 | 0.961052 |
| dj | 0.02704 | 0.071878 | 0.062387 | 0.016761 | 0.038948 |
| wj | 0.124601 | 0.331215 | 0.28748 | 0.077233 | 0.179472 |

3.3确定度计算

通过下式计算某一实测值的确定度，得到该实测值在五个稳定等级下的确定度，其中确定度较大的隶属等级即为多维云模型输出等级。



式中aj为表中wj

4实例验证

为验证该种模型在实际围岩稳定性评价中的可靠性及易用性，按3中方式创建五维度云模型，5个维度即为围岩实测值中的五个指标项，由于总体模型在数据维度上达到6维度，因此无法通过可视化方式表达模型。在此针对样本表中12个样本数据进行计算分级，并对分级结果进行验证，分析其有效性。

评价模型的五个基本指标分别为C1 为岩石质量指标(RQD (%)), C2 为湿抗压强度(Rw/MPa), C3 为完整性系数(kv), C4 为结构面强度系数(k f), C5 为地下水渗水量(W/ L•min/10m)，每个指标分为N1(稳定)，N2(基本稳定)，N3(稳定性差),N4(不稳定),N5(极不稳定)五个等级，其五个标准划分如表所示，样本实测数据如表所示。根据评价标准，已在表中按照式得到云模型Ex，En，He参数，如表所示。根据熵权法计算得到C1到C5各指标所占权重，计算过程如表所示。由式计算确定每个实测样本数据对于各个稳定级别的综合确定度，计算结果见表。

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 样本 | N1 | N2 | N3 | N4 | N5 | 综合评级 | 胡宝清 | 陆兆溱 | 陈守煜 |
| 1 | 0.0000 | 0.0934 | 0.0543 | 0.0000 | 0.0000 | 2 | 2 | 2 | 3 |
| 2 | 0.0000 | 0.0953 | 0.1507 | 0.0000 | 0.0000 | 3 | 3 | 2 | 3->4 |
| 3 | 0.0000 | 0.0229 | 0.5295 | 0.6166 | 0.0000 | 4 | 4 | 4 | 4 |
| 4 | 0.0005 | 0.8519 | 0.0475 | 0.0000 | 0.0000 | 2 | 2 | 2 | 3->2 |
| 5 | 0.0004 | 0.8827 | 0.1334 | 0.0000 | 0.0000 | 2 | 2 | 2 | 3 |
| 6 | 0.0005 | 0.8503 | 0.0505 | 0.0000 | 0.0000 | 2 | 2 | 2 | 3 |
| 7 | 0.0004 | 0.8777 | 0.1351 | 0.0000 | 0.0000 | 2 | 2 | 2 | 3 |
| 8 | 0.0002 | 0.4344 | 0.8303 | 0.1243 | 0.0000 | 3 | 3 | 3 | 4 |
| 9 | 0.0001 | 0.2251 | 0.5068 | 0.4140 | 0.0000 | 3 | 4 | 3 | 4 |
| 10 | 0.0000 | 0.0782 | 0.2086 | 0.0000 | 0.0000 | 3 | 3 | 3 | 4->3 |
| 11 | 0.0000 | 0.0353 | 0.2292 | 0.7489 | 0.0000 | 4 | 4 | 4 | 4 |
| 12 | 0.0000 | 0.0194 | 0.4269 | 0.4808 | 0.0000 | 4 | 4 | 4 | 4 |

由上表可以看出，对比胡宝清的结果，本方法在样本9的评级上存在不同，但是本方法在N3和N4的确定度上很接近，属于介于N3和N4之间，而样本9的结果与陆兆溱的评价结果一致，因此样本9的评级在N3和N4之间是正确的。对比陆兆溱的结果，只有样本2不同，而本方法计算样本2时只有N2和N3对应的确定度显著，因此该样本在N2到N3之间，可以认为本样本结果仍是合理的。

通过上述实际验证过程可知本模型针对该样本的评价结果与现有认可的评价结果并无不合理差距，表明通过熵权法-多维云模型方法来评价围岩稳定性是可行的，而对比文献[汪明武]提出的基于一维云模型的评价模型，其采用权值向量将各指标的确定度线性组合，形成最终确定度，其方法也验证了其可行性，但其方法中的线性组合并不能体现围岩各指标确定度与综合确定度之间的非线性关系，因此本文通过建立多维云模型来探讨确定度非线性组合方式的可行性。

对于本模型采用的熵权法确定指标权重，相比文献[][]中采用的简单关联函数方法确定指标权重，前者针对样本总体，通过统计学的方式计算出各指标在离散程度上表现出的重要性差异，其最终权重与总体样本数量和质量相关，而后者通过对每个样本实测数据与评价标准做对比，计算样本实测值各指标对应的稳定级别的安全威胁大小，通过威胁的大小来确定指标重要性，其每个样本都有自己对应的权重向量。上述方法对比可见，在计算量上前者明显少于后者，并且前者体现出各指标重要性程度是客观确定的，与单一一个样本实测数据相关性不大，而后者方法体现出各指标重要性很大程度取决于自身样本数据的大小。

5结语

本文将熵权法-多维云模型应用在围岩稳定评价中，通过对现有的一维云模型拓展为多维云模型，几倍程度地减少了云模型的数量，并用计算量较小的熵权法取代现有的指标权值确定方法，改变已有方法[汪明武]中的权值参与方式，整体上减少了建模的复杂程度，且仍体现出模型在处理不确定性和模糊性问题中的优势。本方法由于采用熵权法确定指标权重，因此实测样本越多权重越稳定，可将算法集成为学习机，通过不断累积实测数据来不断调整指标权重，使之达到稳定状态。但由于文中样本数仅为12个，因此计算确定的权重有待通过更多样本来更进一步优化。