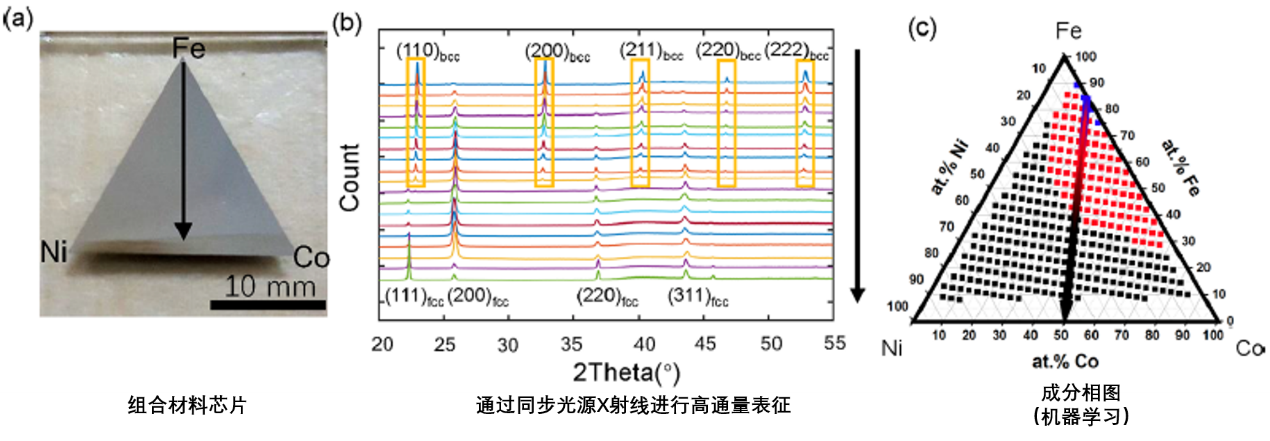
通过组合材料芯片技术快速构建Fe-Co-Ni成分相图



摘要：通过磁控共溅射制备100nm厚的Fe-Co-Ni材料芯片并在500，600，700℃下绝热退火。逐点的成分和结构匹配由同步光源的微束X射线表征。衍射谱的图像按照每秒一张的速率被记录下来。XRD衍射图谱被自动处理，相的识别与归类由层次聚类的算法完成，并用来构建组合成分相图。所构建的相图与ASM合金相图数据库记载的绝热章节内容一致，验证了本文构建相图方法的有效性。

关键词：组合材料芯片、Fe-Co-Ni，X射线衍射，层次聚类，相图

构建相图的传统方法是每次对一个样品进行特征提取和表征，此法耗时，没有系统的流程，在科技飞速发展的现今远不能满足科研人员的需求。自上世纪60年代起，进行过许多匹配成分-相关系的尝试。组合材料技术、特征高通量提取、组合材料库的快速表征展示出加速材料筛选和优化的巨大潜力。Yoo等人利用Fe-NI-Co三元合金系统说明了使用包含连续成分扩散的组合材料芯片薄膜来建立晶体结构和成分之间的复杂关系是可行的。JangHorban等人展示相似类型的样品，Cr-Ni-Re在1100℃下的相图与已发表文献中的相图一致。

通过XRD构建组合材料库的通量很大程度上受限于X射线的流量和射线束斑的大小。同步衍射为高空间分辨率的快速表征提供了一个理想的X射线源。进一步的加速可通过使用带有区域探针的聚焦X射线微束实现，其避免了耗时的角度扫描。经证实，组合材料库的衍射谱可通过1~30s的同步衍射获取。

为了适应高通量衍射实验产生的大量数据，需要自动化相的识别和和聚类的流程。在衍射谱的数据处理工作中用到了机器学习的方法。举个例子，Bunn等人使用XRD、拉曼荧光光谱和成分数据分析Ni-Al薄膜的相形成和氧化物时采用Adaboosting特征学习，一种监督式机器学习方法。该机器方法的方法的一大弊端是需要大量的训练数据，这在实际中通常是不可行的。Long等人使用非负矩阵分解(NMF)，一种非监督式机器学习方法，用于相的匹配以减少对数据量的需求，对基本的物理意义考虑的较少。许多微分方法，如CombiFD、GRENDEL和AgileFD被开发用来确保结论模型携带足够的物理含义。由于XRD峰位的偏移会覆盖较大的成分范围，直到目前，消除峰位偏移仍是对相匹配算法的一大挑战。

聚类是一项根据特定测量将数据分组的技术。在最近的工作中，Iwasaki等人利用取自材料组合数据库的X射线衍射数据结合层次聚类，通过尝试不同的相似度度量方法以比较每种相似度度量方法的有效性。常见的相似度度量方式有以下几种：

1. 皮尔逊相关系数(Pearson Correlation Coefficient)

皮尔逊相关系数一般用于计算两个定距变量间联系的紧密程度，它的取值在[-1，+1]之间。

是x和y的标准偏差

原理：用来反应两个变量线性相关程度的统计量

范围：[-1，+1]，绝对值越大，说明相关性越强，负相关对于推荐的小。

该相似度并不是最好的选择，也不是最坏的选择，只是因为其容易理解，在早期研究中经常被提起。使用Pearson线性相关系数必须假设数据是成对的从正态分布中取得的，并且数据至少在逻辑范畴内必须是等间距的数据。Mahout中，为皮尔逊相关计算提供了一个扩展，通过增加一个枚举类型(Weighting)的参数来使得重叠数也成为计算相似度的影响因子。

1. 欧几里得距离(Euclidean Distance)

最初用于计算欧几里得空间中两个点的距离，假设x，y是n维空间的两个点，它们之间的欧几里得距离是：

可以看出，当n=2时，欧几里得距离就是平面上两个点的距离。当用欧几里得距离表示相似度，一般采用以下公式进行转换：距离越小，相似度越大。

原理：利用欧式距离d定义的相似度S，S=

范围：[0，1]，值越大，说明d越小，距离越近，相似度越高

说明：同皮尔逊相似度一样，该相似度也没有考虑重叠数对结果的影响，同样的，Mahout通过增加一个枚举类型的参数来使得重叠数也成为计算相似度的影响因子。

1. Cosine相似度(Cosine Similarity)

Cosine相似度被广泛用于计算文档数据的相似度：

原理：多维空间两点与所设定的点形成夹角的余弦值。

范围：[-1，1]，值越大，说明夹角越大，两点的距离就越远，相似度就越小。

说明：在数学表达中，如果对两个项的属性进行数据中心化，计算出来的余弦相似度和皮尔逊相似度是一样的，在Mahout中，实现了数据中心化的过程，所以皮尔逊相似度也是数据中心化后的余弦相似度。

1. Tanimoto系数(Tanimoto Coefficient)

Tanimoto系数也称为Jaccard系数，是Cosine相似度的扩展，也多用于计算文档数据的相似度：

原理：是对Cosine相似度的扩展。

范围：[0，1]，完全重叠时为1，无重叠项时为0，越接近1越相似。

说明：适合处理无偏好的打分数据。

1. 曼哈顿距离(Manhattan Distance)

曼哈顿距离公式为：

，为n维曼哈顿空间中的两个对象。

原理：同欧氏距离相似，均用于多维数据空间距离的测量。

范围：[0，1]，同欧氏距离一致，值越小，说明距离值越小，相似度越大。

说明：比欧氏距离计算量小，计算性能相对较高。

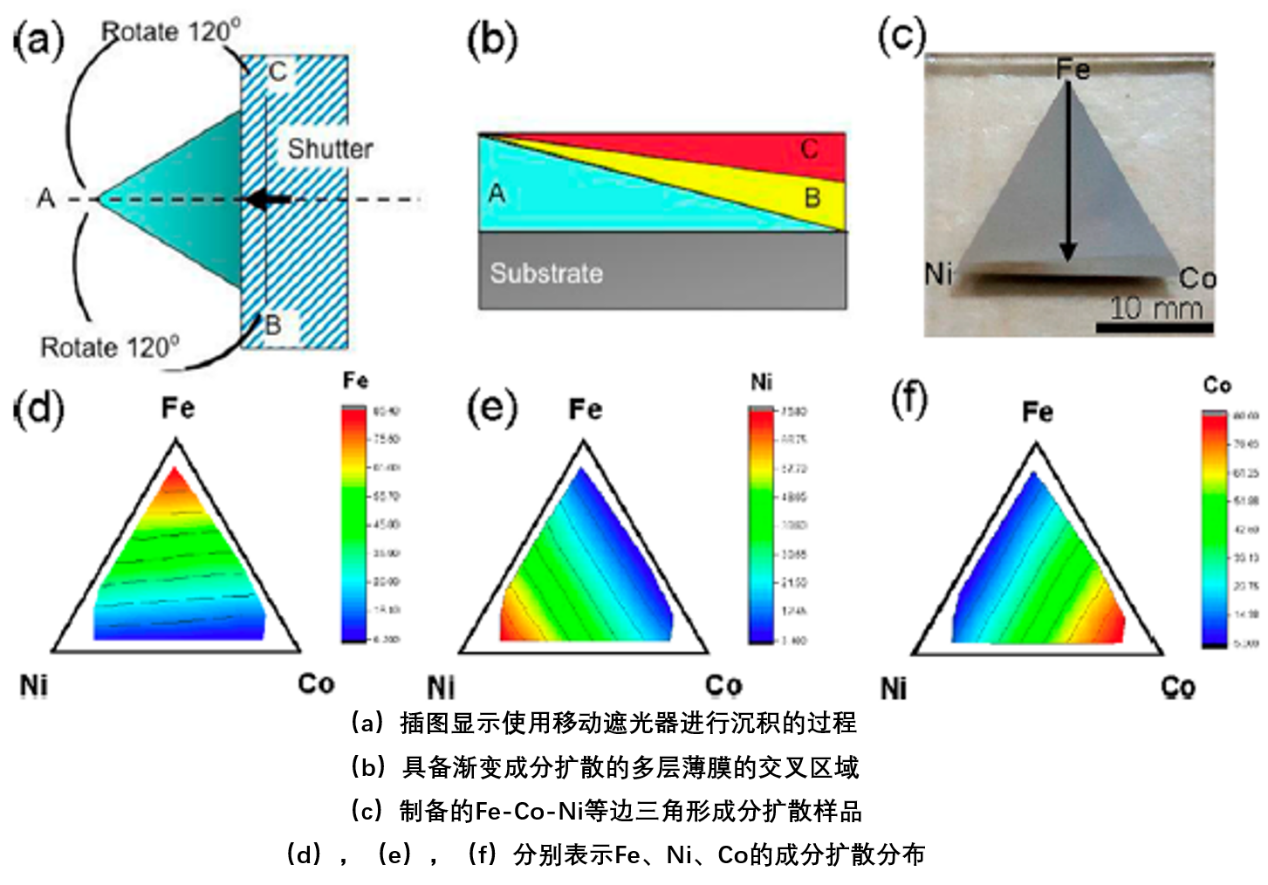
1. 切比雪夫距离公式(Chebyshev Distance)

切比雪夫距离公式为，其中，，为n维曼哈顿空间中的两个对象。

他们发现Cosine、Pearson相关系数和Jensen-Shannon离散度相似度测量技术，在有峰高变化和峰位的随机偏移的情况下，能呈现最好的结果。通过结合吉布斯相规则，Suram等人使用AgileFD-Gibbs算法构建了具备物理含义的V-Mn-Nb氧化物相图。

在本文的工作中，一套快速构建成分相图的系统化的工作流得以建立。工作流中包括组合材料芯片的准备、使用X射线微束衍射和X射线荧光进行成分表征和基于层次聚类的自动化数据分析。为了对工作流进行演示，本文选择记录详实的Fe-Co-Ni三元合金系统。结果成分相图与ASM合金相图的绝热章节部分对比以便确认结果的正确性。

薄膜组合材料芯片能覆盖Fe-Co-Ni三元合金系统的整个成分变化范围。通过使用自行设计的高通量组合离子束沉积系统(HTC-IBD)，芯片被沉积在石英基底上。



芯片上每点的成分由该点的厚度确定

t是样品的厚度，是密度，Z是每个元素的原子质量。

100nm厚的多层薄膜密封在抽真空的石英管中，分别在500℃、600℃、700℃下绝热处理2小时，随后空冷。

各点化学成分的确定是有X射线荧光光谱确定的，因为Fe、Co、Ni三种元素的相对原子质量较近，三种元素的空间分布较为线性化，此现象与实验设计的初衷一致。

两相区中的一个能谱的面心立方和体心立方的晶格常数预测值，分别与Ni的面心立方PDF卡片（no.04-0850）的值0.3523nm及Fe的体心立方PDF卡片（no.06-0696）的值0.2866nm比较后，取为0.356和0.285nm。早期研究表明，Fe和Co混入面心立方的Ni会导致晶格常数的增加，体心立方的Fe的晶格常数随着Co的原子百分比增加至25%一直呈正相关的关系，之后随着Co的原子百分比的增加而递减。因此，合金化后的晶格常数的畸变与合金化的定性趋势是一致的。

为了自动确定相区域，采用层次聚类去给衍射谱分组。Cosine距离被用来计算两条衍射谱之间的相似度，用D表示衍射谱Pm和Pn之间的距离，计算公式如下：

和代表矢量和的第i个元素。

图3a是600℃下热处理的材料芯片所有XRD衍射谱的集合。在聚类的过程中采用了三种策略。首先比较能谱每个衍射角的强度（全局策略），据此在成分空间识别出三个类别。尽管聚类结果并不违反吉布斯相规则，但是两相区的尺寸远大于图2c手绘的结果。如此大的误差主要是由于整个能谱不规则的背底造成的，导致聚类分析中引入较多的噪声。另外，由于相分辨造成的峰的偏移和峰高的改变也被认为是错误聚类的一个主要因素。

其次，