## 通过集群和概率分布确定WLAN位置

我们提出WLAN位置确定技术，联合聚类技术，其使用（1）信号强度概率分布来寻址有噪声的无线信道，以及（2）聚集位置以减少搜索无线电地图的计算成本。与现有技术的当前状态相比，联合聚类技术将计算成本降低超过一个数量级，从而允许在移动客户端上没有集中实现。 来自802.11的iPAQ实施的结果表明，新技术给用户位置在7英尺内超过90％的准确性。

随着无所不在的计算越来越流行，环境感知应用的重要性增加。在转向燃料确定用户位置的需要，利用它该系统可以提供特定位置的信息和服务[6]。多年来的许多系统已经解决了问题，确定和跟踪用户位置。 例如包括GPS [7]，基于广域蜂窝的系统[16]，基于红外的系统[18,3]，基于超声波系统[12]，各种计算机视觉系统[9]cal接触系统[11]和射频（RF）系统[4,20,14,5,13,10,8]。 其中，类的基于RF的系统，其使用基础无线数据网络，工作[4,20,14,5,13,10]，例如802.11，来估计用户位置近来已经获得关注，特别是对于室内应用。 与基于红外线的系统不同，在范围内，基于RF的技术提供更无处不在覆盖和不需要额外的硬件为用户确定位置，从而增强无线数据网络的价值。

基于RF的系统需要处理无线信道的嘈杂特性。 这些特性造成在线阶段测量的样本从存储在无线电地图中的那些信息中显著偏离，限制了系统的准确性。 此外，为了保留用户隐私，并使定位系统可扩展，位置确定代码应该在移动单元上运行。 由于移动设备是能量约束的，位置确定系统中减少计算要求相当重要。

在本文中，我们提出一个精确和可扩展的系统 - 用于确定具有低计算 - 802.11无线局域网（WLAN）框架。 该系统有两个主要特点：（1）它使用概率分布提高准确性和处理无线信道的噪声性质。 （2）它使用地图位置聚类，以减少计算要求。我们称之为我们的技术联合聚类（JC）技术。

我们在跨度为20000平方英尺的室内空间下评估该系统。 结果表明联合聚类技术在计算非常低的情况下，在7英尺以内给用户定位的准确度超过90％的精度。

基于无线电地图的技术可以分类为两大类：确定性技术[4，14]和基于分布的技术[20，5，13，10]。 我们的工作在于在第二类。 然而，以前的系统考虑到确定定位算法的复杂度。通过介绍把减少无限点地图聚合定位来作为减少定位技术的计算量和提高系统可扩展性的入门， 我们的工作是独一无二的。

本文的其余部分组织如下。 在Sec-第2节，我们描述无线的嘈杂特性渠道。 第3节介绍无线电地图的细节结构和位置估计与联合聚类技术。 在第4节，我们描述的评价技术在室内空间和获得的结果。 Fi-最后，第5节结束了论文并给出了指导未来的工作。

1. 无线信道特性

在本节中，我们描述我们的采样过程和无线信道的噪声特性，这使得WLAN位置确定的问题是一个具有挑战性的问题。

2.1。 采样过程

所有WLAN位置确定系统所需的关键功能是信号强度采样。 为了本文的目的，我们使用了Lucent Orinoco silver NIC，支持高达11 Mbit / s的数据速率。

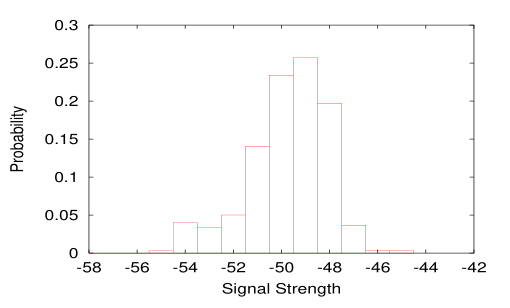
我们修改了用于Linux的Lucent Wavelan驱动程序，以便使用主动扫描返回从NIC范围内的所有接入点接收的信标帧的信号强度; 我们的驱动程序是第一个在Linux下支持此功能的。 我们还开发了一种无线API，可与支持无线扩展的任何设备驱动程序连接。 设备驱动程序和无线API已可供公众下载，并已被其他人用于无线研究。

2.2。 嘈杂特性

IEEE 802.11b标准在2.4GHz使用无许可频带中的无线电频率。 虽然这个许可自由有助于在基于802.11b的网络的广泛使用，它也有它的问题。 在2.4GHz频段，蓝牙设备，2.4 GHz无绳电话，微波炉等设备可能成为干扰源。 此外，2.4GHz是水的共振频率，人体可以吸收RF信号。

多径衰落[15，19]是另一种常见的在RF波传播中的现象。 发送信号可以通过不同的路径到达接收器，每个具有其自身的幅度和相位。 这些不同的组件组合并产生传输的失真版本信号。 此外，环境条件的变化，例如温度或湿度，影响接收的信号强度。

图1给出了一个典型的例子，从接入点接收的信号强度的柱状图。 人们在环境，门打开和关闭等环境变化可以解释图中所示的这种时间变化。

图1.一个接入点的信号强度的直方图例子

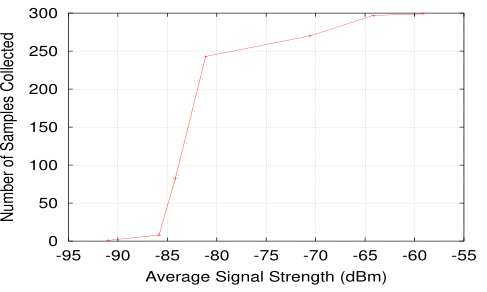


图2.平均信号之间的关系,接入点的强度和从其接收的样本的百分比。

我们还进行了一个实验来测试相同位置具有不同的平均信号强度接入点的表现。 在这个实验中，我们对信号每个接入点以一个样本的速率，采样接收强度每秒。 图2示出了每五分钟采样一次的信号中从接入点接收的平均信号强度百分比。 该图显示了从接入点收集的样本的平均信号功率是单调递增的。 假设噪声电平恒定，信号强度越强，信噪比越高，更可能成为802.11b卡可识别的的数据包。 在-81 dBm的急剧下降可以通过注意接收机的灵敏度来解释我们使用的卡是-82 dBm（在11 Mbps）。

总之，图1和图2突出的无线信道特征：

•在固定位置，从接入点接收的信号强度随时间变化。

•覆盖位置的接入点的数量随时间变化。

下一节介绍联合聚类技术以解决这些噪声特性。

1. 联合聚类技术

我们将集群定义为共享一组公共接入点的一组位置。 我们称这个公共集合的接入点为集群密钥。 联合聚类技术使用不同接入点的信号强度的联合概率分布来找到给定观测的信号强度值的最可能的用户位置。另外，它使用聚类来减少计算开销。 因此，联合聚类技术的操作可以分为两个阶段：（1）离线阶段，我们执行联合分布估计和位置聚类和（2）位置确定阶段，其中我们运行位置确定技术来推断用户位置。 下面，我们将描述两个阶段更多细节。

3.1。 离线训练阶段

在离线阶段，我们执行两个任务：联合概率分布估计和位置聚类。

3.1.1。 估计联合信号强度分布

在训练位置集合中的每个位置处，我们存储用于在该位置处的接入点的联合概率分布的模型。 因此，我们的无线电地图存储为用于联合概率分布的模型的集合。可以进一步估计关节分布的问题可分为三个子问题：

1.连接值，连接点的大小分配？

2.哪些k个接入点，来自接入点集合覆盖某一位置，选择包括在内联合分布？

3.如何估计所选择的k个接入点之间的联合分布？

确定k的最佳值。在确定最好k的值，我们需要考虑2个因素：（1）随着k增加，估计联合概率分布（子问题3）的过程变得更复杂，以及（2）我们需要一个k的值，使得所有位置在大多数时间由至少k个接入点覆盖。第二个因素很重要，因为给定位置的接入点数量随时间变化（图2）。 第二个因素减少了接入点数量变化的影响，因此应该导致更好的精度。 参数k的典型值可在第4节中找到。

选择k个接入点。 如果覆盖某个位置的接入点数量随时间变化，我们应该选择哪些接入点？ 直观地，我们应该选择在样本中大部分时间出现的接入点。 图2建议我们应该选择使用在每个位置具有最大信号强度值的k个接入点。

总之，对于给定位置l，我们选择覆盖该位置的k个最强接入点。

估计联合概率分布。 联合概率分布可以以不同的方式用不同的精度水平来估计。 该问题可以表示为：给定k个接入点AP 1，...，AP k，我们想要估计P（AP 1 = s 1，...，AP k = sk），其中si是来自 AP i。 估计联合分布的一个好方法是使用最大似然估计方法，其将联合概率估计为：

即，在整个训练集中出现的信号强度值数组（s 1，s 2，...，s k）除以训练集的大小的次数。

这种方法的问题是，它需要一个大的训练集，以获得联合分布的良好估计，所需的大小随k呈指数增长。 因此，该方法只能用于小的k值，这可能影响技术精度。

因为我们的目标是使用一种提供良好精度的方法，同时需要合理数量的训练数据和计算能力，我们选择假设接入点是独立的。 这个假设对于良好设计的802.11网络是合理的，其中每个接入点在非重叠信道上运行。 此外，独立性是第3.2节中讨论的常见的贝叶斯假设。 因此，估计联合概率分布的问题随后成为估计边际概率分布的问题：



对于给定位置，可以使用在该位置处的接入点AP i的归一化直方图来估计P（AP i = s i）。 图1给出了来自接入点的信号强度归一化直方图的典型示例。

3.1.2。 位置聚类

为了减少计算开销，我们根据覆盖的接入点将位置分组成簇位置。 问题可以说如下：给定一个位置l，我们要确定l属于哪个集群。

最明显的做聚类的方法是根据覆盖它们的接入点对位置进行分组。 即：两个位置l 1和l 2被放置在相同的群集中覆盖这些位置的接入点组是相同的。然而，这种用于聚类的方法在应用于真实环境中时具有问题。 如图2所示，接入点可能从一些采样丢失，因此，使用覆盖用于聚类的位置的接入点的整个集合可能由于丢失的接入点而无法找到正确的群集。

用于代替使用覆盖位置的接入点集合，我们使用该集合的仅包含q个元素的子集，并且问题变为：给定数字q，我们希望将共享q个接入点的所有位置放置在一个集群中。 因此，我们有2个子问题：1.如何确定q的值？

2.哪些q个接入点可以选择用于聚类？

对于第一个子问题，我们需要选择q，使得所有位置在大多数时间由至少q个接入点覆盖。 这个因素是重要的，因为类似的原因，如在讨论参数k的值的选择。 这表明q的值应小于或等于覆盖无线电地图中任何位置的接入点的最小数量。 此外，我们需要一个q值，它在集群之间均匀分布位置，以减少所需的计算。 q的典型值在第4节中给出。

对于子问题2，我们选择使用在每个位置具有最大信号强度值的q个接入点，再次是由于与前面部分中类似的原因。

在数据分析期间，我们发现在某些位置，当来自这些接入点的信号强度值彼此接近时，具有最大信号强度值的接入点的顺序改变。 因此，我们选择将q个接入点视为一个集合，而不是一个有序的元组。

总之，对于给定位置l，我们使用覆盖该位置的q个最强接入点的集合来确定其所属的集群。 因此，集群密钥是用于对该集群中的位置进行分组的q个接入点的集合。

我们这里要强调的是，参数k（联合分布的维度）和q（在聚类中使用的接入点的数量）的值是独立的。

3.2。 在线定位阶段

在位置确定阶段期间发生的一般想法如下：我们从在未知位置的一些接入点获得样本。 我们使用q个最强的接入点来确定一个群集在最可能的位置内搜索。 然后，我们使用贝叶定理来估计在给定观察样本和在离线阶段期间建立的无线电图的集群中每个位置的概率。 最可能的位置被报告为估计的用户位置。 上述算法工作假设理想的无线信道。 然而，在实际环境中，我们需要解决两个问题：

1.在一个位置处的测试样本中的接入点的数量可以小于在聚类中使用的接入点的数量q。

2.在一个位置处的测试样本中的接入点的数量可以小于联合分布的维度k。

为了解决第一个问题，我们搜索其密钥具有{AP 1，AP 2}作为子集的所有集群。 我们使用这些集群中所有位置的并集作为我们的目标位置集。如果测试样本中的接入点数量大于或等于q，则目标位置集合会减少到一个集群中的位置。

对于第二个相似的问题，我们使用相同的方法来通过将联合分布的维数减小到min（k，测试样本中的接入点的数目）来解决它。

唯一需要解释的是如何使用贝叶定理计算给定观测向量的目标位置中最可能的位置

¯

S =（s 1，...，s k）。 我们想找到使得P（1 / S）最大化的l。 即我们想要的

使用贝叶斯定理，这可以重写为：

由于对于所有l，P（S）是常数，所以可以将公式4重写为：

可以基于如果用户在给定位置处的事实，他更有可能在将来处于相邻位置的事实，从用户简档确定P（1）。 如果用户简档信息是未知的或未使用的，则我们可以假定所有位置同样可能，并且项P（l）可以从最大化过程中被计算出。 等式6变为：

如3.1节所述，剩余项通过以下公式计算：

根据存储在无线电地图中的信号强度分布来估计P（s i / l）

可以用同时使用n个样本的序列代替使用来自每个接入点的一个信号强度样本来估计用户位置。 假设样本的独立性，则对于给定位置，通过乘以每个样本的概率来获得样本序列的概率。

使用观察序列，而不是一个样品，具有以下优点：

•由于我们从更多样品中获得更多信息，因此应提高精度。 细节在第4节。

•这也有助于聚类：随着序列长度增加，接入点将丢失的概率降低。 如果我们从接入点获得样本的概率是p，则我们在n个样本的序列中获得至少一个样本的概率是1-（1-p）n。 例如，如果p = 0.8和n = 2，则获得样本的概率从0.8增加到0.96。

下一节讨论“联合丛书”技术。

3.3。 讨论

算法的存储器需求是有限的。如果每个位置的接入点的平均数量是4并且每个接入点的平均范围是11个不同的值，则对于每个位置，我们需要存储11 \* 4个参数，对应于直方图 每个接入点，这是一个小数字。 我们可以通过连续分布来近似直方图，例如， 对数正态分布，并且仅保存每个接入点的分布的均值和方差。 然而，这种近似会影响系统的精度，并且节省存储器要求并不合理。

联合聚类技术使用的聚类技术减少了搜索空间，从而导致计算成本的降低。 此外，使用聚类有助于将系统缩放到更大的覆盖区域。

1. 实验评价

在本节中，我们讨论实验测试台，并评估联合聚类技术的性能与之前在WLAN位置确定领域的工作进行比较。 所有系统都在相同的环境中实施以进行公平比较。

4.1。 实验测试床

我们在计算机科学部大楼四楼的南翼进行了我们的实验。地板的布局如图3所示。机翼的尺寸为224英尺乘85.1英尺。 联合聚类技术在计算机科学部无线网络中进行了测试。

为了建立无线电地图，我们在走廊上的电台地图位置在一个网格上的单元格放置5英尺（走廊的宽度为5英尺）。 我们在走廊上总共有110个地点。 平均来说，每个位置由4个接入点覆盖。

使用我们开发的设备驱动程序和API，我们在每个位置收集300个样本，每秒一个样本，并且我们使用它使用先前描述的方法来估计每个位置处的每个接入点（无线电地图）的分布。 训练数据的大小对性能的影响在第4.4节中讨论。 为了测试系统的性能，我们使用了在不同的日子，一天的时间和不同于训练集的人收集的独立测试集。

4.2。 绩效评估

•精度：此度量定义为技术在一定距离内给出正确位置估计的时间百分比。

•每个位置估计的操作数：此度量定义为对单个位置估计执行的操作（乘法）的总数。 这在最小化计算时间中是重要的，但是在最小化功耗方面更重要。

4.3。 参数对性能的影响

联合聚类技术具有两个控制参数。 在本节中，我们研究这些参数的影响，特别是k（联合分布的维度）和q（在聚类中使用的接入点的数量）对其性能的影响。

我们首先展示改变q对聚类过程的影响。对于这个实验，我们将q的值从1改为4，并计算簇数，每个簇的平均大小和簇大小的标准偏差。这在图4中示出。从图中可以看出，随着q增加，簇的数目增加，并且每个簇的平均尺寸减小，直到在q = 2处达到饱和点。对于标准偏差，群集的大小减小，直到我们达到最小值，在q = 3，并且其再次增加。由于群集的最小值更加均匀，这是期望的属性。 q = 3处的最小值可以通过注意当q从1增加到3时，由于添加新的接入点而将更多的位置区分为不同的簇来解释。当q增加超过3，即q = 4时，不同的位置开始共享相同的4个接入点，特别是对于彼此靠近的位置（回想在我们的实验中每个位置的接入点的平均数量是4），因此每个簇的位置的数量开始偏离在簇之间的均匀性，导致增加的标准偏差。

图5和图6示出了参数q和k一起对性能的影响。 从图中我们看到，随着维度k增加，精度增加，因为由于添加接入点而具有更多的信息，并且由于相同的原因，每个位置估计所需的操作的数量增加。 随着在聚类（q）中使用的接入点的数量增加，每个群集的元素的数量减少，导致每个位置估计的精确度增加和操作数量减少。

对于本文的其余部分，我们选择将参数的值作为q = 3和k = 4，因为这些值导致我们的实验的联合聚类算法的最佳性能。

4.4。 结果

在本节中，我们展示了联合聚类技术的性能评估。 我们还将其性能与雷达系统的性能进行比较[4]，在同一个测试台中实现，作为参考点。

图7示出了两种技术的误差距离的CDF。 联合聚类技术在7英尺内提供超过90％的精度，而在雷达系统中相同距离范围的精度为38％。

在图8中示出了在每个采样的平均操作次数方面的两种技术之间的比较。该图示出了使用聚类将每个位置估计的操作的平均数量减少超过一个数量级。

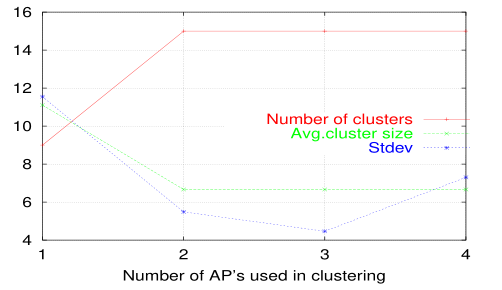
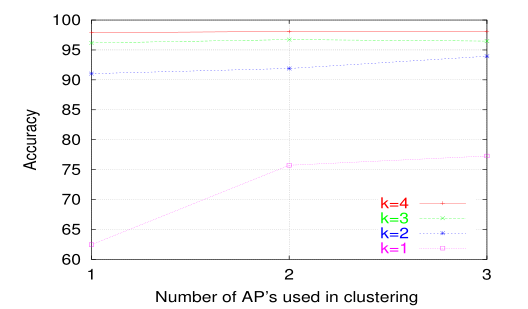
图4. q对聚类过程的影响。

图5.参数q和k对精度（14英尺内）的影响。

4.5。 测试序列长度的影响