Hmm 应用

Hmm 的应用主要分为 4 类。计量经济用法，将隐状态作为预测自变量法，利用相似度预测法，一种语音识别（分类）法。

计量经济用法（文献 1,14）

1) 基本用法

假设我们的数据是日数据，第 t 日的数据记为 $X_t=\left(X_{1t},X_{2t},X_{3t}\right)',t=1,\ldots T$ ，其中，X_{1t},X_{2t},X_{3t} 分别表示 t 日的收盘价，成交量，相对前一日收盘价的涨跌幅（这些变量可以是其他变量，但是必须包含收盘价或相对于前一日收盘价的涨跌幅），T 表示总数据长度。我们将数据分为样本内和样本外，样本内数据为 $t=1,\ldots,T_{in}$ ，样本外数据为 $t=T_{in}+1,\ldots,T$ 。

我们假设时间序列 $X_t,t=1,\ldots,T$ 服从 hmm 模型，该模型隐状态为 N 个，发射分布（即条件在 t 时刻的状态上，X_t 的概率分布）为正太分布，该模型中待估计（待学习）的参数记为 $\lambda=\left(\pi,A,B\right)$ 。我们使用样本内数据训练，得到 hmm 模型的参数。

然后做样本外预测。具体为，对于 $t=T_{in}+1,\ldots,T$ ，我们可以得到 $P\left(X_t\middle|X_1,\ldots,X_{t-1}\right)$ 或 $E\left(X_t\middle|X_1,\ldots,X_{t-1}\right)$。可以看到，该预测是条件在 $1,\ldots t-1$ 上的所以变量的，随着时间的增加，这会增加计算复杂度，所以，我们可以计算 $P\left(X_t\middle|X_{t-w},\ldots,X_{t-1}\right)$ 或 $E\left(X_t\middle|X_{t-w},\ldots,X_{t-1}\right)$，即该预测是条件在最近 w 个时刻的信息上的，这是上述完整信息的一种近似。

2) 用法分析

该用法的本质是，对 X_t 与 $X_1,\ldots,X_{t-1}$ 之间的关系建立了模型，类似于自回归模型，自回归模型是假设，

$$X_t=\mu+C_1X_{t-1}+C_2X_{t-2}+,\ldots,C_lX_{t-l}+\varepsilon_t \tag{1.1}$$

即，X_t 是过去 l 个滞后项的线性函数再加上一个随机扰动项。而 hmm 模型是假设，X_t 受 t 时刻隐状态 S_t 决定（发射分布），而 S_t 又由 S_{t-1} 决定（马尔科夫状态转移矩阵），同样，X_{t-1} 又由 S_{t-1} 决定，而 S_{t-1} 又由 S_{t-2} 决定，…，从而建立起了 X_t 与 $X_1,\ldots,X_{t-1}$ 之间的联系。具体为，

$$
\begin{aligned}
&P\left(X_t = x_t \middle| X_1 = x_1, \ldots, X_{t-1} = x_{t-1}, S_1 = s_1, \ldots, S_{t-1}, S_t = s_t\right) \\
&= P\left(X_t = x_t \middle| S_t = s_t\right) \\
&= N\left(x_t; \mu_i, \sigma_i\right)
\end{aligned}
\tag{1.2}
$$

$$
\begin{aligned}
&P\left(S_t = s_t \middle| X_1 = x_1, \ldots, X_{t-1} = x_{t-1}, S_1 = s_1, \ldots, S_{t-1} = s_{t-1}\right) \\
&= P\left(S_t = s_t \middle| S_{t-1} = s_{t-1}\right)
\end{aligned}
\tag{1.3}
$$

其中，大写字母为随机变量，小写字母为取值， $N\left(x; \mu_i, \sigma_i\right)$ 为正太分布， $s_t, t = 1, \ldots T$ 的取值范围为 $\left\{1, \ldots, N\right\}$ 。(1.2)式也可以写成

$$
\begin{aligned}
&X_t = \mu_{S_t} + \varepsilon_{tS_t} \\
&\varepsilon_{tS_t} \sim N\left(0, \sigma_{S_t}\right)
\end{aligned}
\tag{1.4}
$$
。

上述(1.2)和(1.4)式假设 X_t 条件在过去所有观测和隐状态上的概率分布只与当期状态 S_t 的取值有关，该假设的更一般形式是

$$
\begin{aligned}
&P\left(X_t = x_t \middle| X_1 = x_1, \ldots, X_{t-1} = x_{t-1}, S_1 = s_1, \ldots, S_{t-1}, S_t = s_t\right) \\
&= P\left(X_t = x_t \middle| X_{t-l} = x_1, \ldots, X_{t-1} = x_{t-1}, S_t = s_t\right) \\
&= N\left(x_t; \mu_i + C_{i1} x_{t-1} +, \ldots, + C_{il} x_{t-l}, \sigma_i\right)
\end{aligned}
\tag{1.5}
$$

即， X_t 条件在过去所有观测和隐状态上的概率分布不仅与当期状态 S_t 的取值有关，还与滞后 l 阶的 X 有关。(1.5)写成

$$
\begin{aligned}
&X_t = \mu_{S_t} + C_{S_t 1} X_{t-1} +, \ldots, + C_{S_t l} X_{t-l} + \varepsilon_{tS_t} \\
&\varepsilon_{tS_t} \sim N\left(0, \sigma_{S_t}\right)
\end{aligned}
\tag{1.6}
$$

可以看到，(1.6)跟(1.1)很像，都是自回归形式，只不过(1.6)的系数是时变的，具体由 t 时刻的隐状态取值决定，隐状态的每个取值对应一套自回归系数。所以该模型又称为 regime-switching 模型。

该用法是对数据生成过程做了假设，即假设我们观察到的数据由背后的隐藏状态决定，比如（熊市状态，牛市状态，震荡状态）或者（市场有效状态，市场非有效状态），该模型好不好，跟该假设是否正确有很大关系。

将隐状态作为预测自变量（文献 2~7）

1) 基本用法

该做法的基本思路是，在每一时刻，利用最近一段时间的数据求出当前时刻的隐藏状态，将当前时刻的隐藏状态作为预测自变量来预测接下来一段时间的涨跌。

具体为，假设我们的数据是日数据，第 t 日的数据记为 $X_t = \left(X_{1t}, X_{2t}, X_{3t}\right)', t = 1, \ldots T$，其中，$X_{1t}, X_{2t}, X_{3t}$ 分别表示 t 日的收盘价，成交量，相对前一日收盘价的涨跌幅（这些变量理论上可以是任何变量），T 表示总数据长度。我们将数据分为样本内和样本外，样本内数据为 $t = 1, \ldots, T_{in}$，样本外数据为 $t = T_{in} + 1, \ldots, T$。

在第 t 日，$t = P_1 + 1, P_1 + 2, \ldots, T_{in}$，使用前 P_1 日数据（$t - P_1, \ldots, t - 1$），估计一个 hmm 模型（该示例是每一天使用最近的数据估计一个 hmm 模型，实际应用中可以每隔几天估计一个 hmm 模型，或者只是最开始估计一个 hmm 模型），然后，使用最近 P_2（$P_2 \le P_1$）日数据（$t - P_2 + 1, \ldots, t$）得到（维特比算法）第 t 日的隐藏状态概率向量。这样，我们在样本内，就得到这样一个样本集合，其中每个样本点为：自变量是 t 日（$t = P_1 + 1, P_1 + 2, \ldots, T_{in}$）的隐藏状态或隐藏状态概率向量，因变量是下一日（或下几日）涨跌幅。我们可以在样本内训练一个机器学习模型，用 t 日隐藏状态来预测 t+1 日涨跌幅。

在样本外，第 t 日，$t = T_{in} + 1, \ldots, T$，使用前 P_1 日数据（$t - P_1, \ldots, t - 1$），估计一个 hmm 模型，然后，使用最近 P_2（$P_2 \le P_1$）日数据（$t - P_2 + 1, \ldots, t$）得到第 t 日的隐藏状态概率向量。然后使用样本内训练好的机器学习模型，来预测 t+1 日涨跌幅。

2) 用法分析

该做法从理论完备性上和实操复杂度上都不如计量经济用法好。

理论完备性上，该做法既然假设变量 X_t 服从 hmm，那么只要在这些变量中包含收盘价或相对于前一日收盘价的涨跌幅，那么就可以直接按照计量经济法来做预测（不管其假设对不对，起码后续推导没问题），而该做法绕了一大圈。之所以绕一大圈，主要原因是在使用 hmm 做量化投资时，基本上都是假设隐藏状态是熊市、牛市、震荡这 3 种状态，然后通过观测数据 X_t 来识别背后的状态究竟是这三种状态的哪一种。值得注意的是（a）就算数据背后的隐藏状态是熊市、牛市、震荡这种划分，使用计量经济法也比该法更符合 hmm 的逻辑。即当日隐藏状态决定次日隐藏状态，次日隐藏状态决定次日观测变量。（b）观测数据背后的隐藏状态不一定非要按照熊市、牛市、震荡这种划分，也可以按照其他方式划分，比如知情交易主导，噪声交易主导这种划分。在这种知情-噪声状态划分下，我们可以做下面的应用：在这种划分下我们可以识别出最近几日哪些是知情交易日，哪些是非知情交易日，然后将知情交易日的带方向的成交量加起来，做个标准化，把它作为因子，用来表示最近的带方向的知情交易程度，而带方向知情交易程度对接下来一段时间的涨跌有预测作用是一个已知的比较合理的逻辑。所以这种利用 hmm 的方式比将隐藏状态作为自变量的方法更有逻辑一些。

实操复杂度上从上面的操作步骤来看，明显该方法要比计量方法步骤多。

利用相似度预测法（文献 8~10）

1) 基本用法

该用法与时间序列分类相同，即观察到一段时间序列（例如最近 5 日市场数据），想要预测接下来一日（或几日）是涨是跌，那就去计算历史上的这种 5 日数据的片段与该片段的相似度，将相似度最高的片段次日的涨跌幅用来预测当前片段次日的涨跌幅。相似度的衡量有很多，最简单的是欧式距离，而 hmm 的用法是，假设当前片段和历史片段都服从 hmm，计算当前片段的似然值，也计算历史片段的最大似然值（hmm 模型参数的估计使用当前片段之前的数据估计），这两个似然值越接近，相似度越高。

具体为，假设我们的数据是日数据，第 t 日的数据记为 $X_t=\left(X_{1t},X_{2t},X_{3t}\right)',t=1,\ldots T$，其中，$X_{1t},X_{2t},X_{3t}$ 分别表示t 日的收盘价，成交量，相对前一日收盘价的涨跌幅（这些变量理论上可以是任何变量），T 表示总数据长度。整个数据都是样本外数据。

在第t 日，$t=P_1+1,P_1+2,\ldots,T$，使用前 P_1 日数据（$t-P_1,\ldots,t-1$），估计一个 hmm 模型（λ）。记时间序列片段长度为 P_2，然后计算观察到 $X_{t-P_2+1},\ldots,X_t$ 的概率 $L\left(X_{t-P_2+1},\ldots,X_t,\lambda\right)$，即当前似然值。然后对于历史样本点 $s=t-P_1,\ldots,t-1$，计算观察到 $X_{s-P_2+1},\ldots,X_s$ 的概率 $L\left(X_{s-P_2+1},\ldots,X_s,\lambda\right)$，选取历史样本点钟似然值与当前似然值最接近的样本点，将这些历史样本点次日涨跌幅的平均作为对当前样本点次日涨跌幅的预测。

2) 用法分析

该用法基本逻辑没问题。一个问题是为什么要用这种方式来衡量两个时间序列片段的相似性。该相似性指标逻辑有疑问。如果假设数据服从均匀分布，那么历史片段和当前片段的似然值之差并不能衡量相似性，因为任何历史片段与当前片段具有相同的似然值。类似地，如果数据服从正太分布，历史片段与当前片段的似然值之差并不能衡量相似性，因为如果历史片段是正太分布的 5%分位数，而当前片段是正态分布的 95%分位数，那么这两个片段的似然值是一样的，但是其实两个片段的取值差异很大。所以，用似然值衡量相似性似乎并不直观。

一种语音识别用法（文献 11~12）

1) 基本用法

由于该用法源自机器学习分类算法的一种方，即 LDA，所以为了理解该用法，我们回顾一下 LDA 的基本原理。LDA 原理如下图：

2) LDA（linear discriminant analysis）

当类型大于等于 3 个的时候，简单的线性回归分类效果差（原因没有解释，只是举例说明）。

记，某个观测属于第k类的先验概率为π_k，某个观测条件在第k类上x出现的概率为$f_k(x)$。则，根据 bayes 公式，在观察到x后，该观测属于第k类的概率为

$P(G=k|X=x)=\dfrac{\pi_k f_k(x)}{\sum_{l=1}^{K}\pi_l f_l(x)}$ （7）。判断某个观测属于哪一类，就看哪一类的$P(G=k|X=x)$最大。许多方法就是对$f_k(x)$建模：

a) 正态分布，会得到 LDA 和 Q（quadratic）DA;
b) 混合正太分布；
c) 非参数的密度函数估计；
d) 原始 bayes，即假设x的各个成分之间独立。

图 1 LDA 原理

其中，x 是特征，k 是类型。常见情况下x 是个向量，比如在划分某个西瓜是否成熟时，x 为向量$(西瓜颜色，西瓜长度，西瓜宽度)$；而在时间序列分类问题中，x 是一个 sequence、，为$(x_1,\ldots,x_\tau)$ ，其中τ 为序列长度。

在一类语音识别问题中，该问题是给定一段音频$x=(x_1,\ldots,x_\tau)$，决定该音频对应哪一个词语k 。该问题的解决，就是按照上图中公式（7）来解决的。只不过，假设，$f_k(x)$ 服从 hmm，即第k类（第k 个词）的观测序列x 服从参数为λ_k 的 hmm $f_k(x,\lambda_k)$。另外，假设不同类型的先验概率相同，即$\pi_k=\dfrac{1}{K}$。所以，对于一个新的样本x ，计算使得（7）式最大的类型，也就是计算使得$f_k(x,\lambda_k)$最大的类型k ，即是对这一样本的分类。

而$f_k(x,\lambda_k)$的参数$\lambda_k, k=1,\ldots,K$的估计为：记类型为k的样本N_k个，第k类的第i 个样本的类型记为i ，则N个观测的联合密度（似然函数）为

$$\prod_{k=1}^{K}\prod_{i=1}^{N_k} f_k(x^i,\lambda_k) \tag{1.7}$$

我们对该式取对数，然后将其最大化，就得到对λ_k的估计。其中，对λ_k的估计相当于最大化

$$\sum_{i=1}^{N_k} \log f_k\left(x^i, \lambda_k\right)$$

(1.8)。

注意，经典的 hmm 估计（Baum-Welch 算法）是假设只有一个 sequence，即(1.8)假设项中的一项，而上述问题是有 N_k 个 sequence，最大化(1.8)式对应于多序列的 Baum-Welch 算法（文献 13）。

同样，在我们使用一段时间序列预测涨跌类型时，自变量是一段时间序列 $x=\left(x_1,\ldots,x_\tau\right)$，即 τ 日的变量，我们要将其分为 K=2（接下来一段时间涨、跌）或 K=3 类（接下来一段时间涨、跌、平）。

2) 用法分析

该做法理论推导没问题。唯一的疑问是，对于不同的类型(接下来一段时间涨、跌)，假设其观测 x 服从不同的 hmm 的用意是什么。为此，我们查看了一下该方法用在语音识别上的动机（文献 13）。假设，音频观测长度为 τ=10。其动机是，对于一个词“你好”，我们对其对应的音频生成规则可以做如下下表假设：

表 1 “你好”的隐状态和观测

音素	n	n	i	i	h	a	o	o	o	o
观测	x1	x2								x10

其中，将音素作为 hmm 中的隐状态，观测为 $x=\left(x1,\ldots,x10\right)$。类似地，对于另外一个词“我好”,我们对其对应的音频生成规则可以做如下下表假设:

音素	w	w	w	o	h	h	a	o	o	o
观测	x1	x2								x10

。很明显，确实，不同的词，其后面有不同的音素结构 λ（概率转移矩阵，发射矩阵，初始概率）。

所以，这种假设（不同的类型假设其观测 x 服从不同的 hmm）对于语音识别是一种合理的假设。对于股价涨跌分类合不合理，由于目前想不出类似于语音分类中音素这样有明确的含义的的东西作为隐状态，所以不好说；但是我们假设有这种东西存在，我们只是不知道它是什么，所以可以尝试。

参考文献

【1】I. Róbert Sipos ∗, Attila Ceffer, Gábor Horváth and János Levendovszky，Parallel MCMC sampling of AR-HMMs for prediction based option trading

【2】东方证券-《衍生品研究系列之四》：基于单特征因子的隐马尔科夫模型在商品期货上的应用

【3】中信期货-量化专题报告：隐马尔可夫模型商品期货应用初探

【4】东北证券-金融工程：HMM 指数择时研究之实战篇

【5】东北证券-HMM 指数择时研究之理论篇

【6】东证期货-【金融期货】包含基本面信息的 HMM 投机策略
【7】量价特征因子：基于 HMM 的多空策略

【8】兴业证券-股指期货交易策略系列报告之三：基于隐马尔科夫链的交易策略
【9】Nguyet Nguyen，Int. J. Financial Stud. 2018, 6, 36，Hidden Markov Model for Stock Trading
【10】Md. Rafiul Hassan and Baikunth Nath，Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications，Stock Market Forecasting Using Hidden Markov Model: A New Approach

【11】广发证券-投资策略-探寻西蒙斯投资之道：基于 HMM 模型的周择时策略研究
【12】广发证券-金融工程专题报告：再探西蒙斯投资之道，基于隐马尔科夫模型的选股策略研究
【13】Larence Rabiner，proceedings of the IEEE，Vol 77,No 22，Feb 1989，A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition
【14】James D. Hamilton， 1994， Time Series Analysis， Chapter 22.