根据上述元素，我们将隐马尔可夫模型下的中文分词问题表述为如下形式：

当我们观测到句子$w_1, w_2 \ldots w_n$，其中w_i为第 i 个汉字，我们希望找到相应的标签序列$s_1, s_2 \ldots s_n$，其中s_i为w_i对应的标签（BMES中的一种），使得$P(s_1, s_2 \ldots s_n | w_1, w_2 \ldots w_n)$概率最大。

为求解这一目标函数，需要隐马尔可夫模型的两个基本假设：

- 齐次假设：当前隐藏状态只与上一个状态有关系
  - 即$P(s_i|s_{i-1}, s_{i-2} \ldots s_1) = P(s_i|s_{i-1})$
- 观测独立性假设：观测值之间互相独立的，只与生成它的状态有关系
  - 即$P(w_1, w_2 \ldots w_n|s_1, s_2 \ldots s_n) = P(w_1|s_1)P(w_2|s_2) \ldots P(w_n|s_n)$

基于目标函数和两个基本假设，基于隐马模型的中文分词问题转化为：

我们希望找到相应的标签序列$s_1, s_2 \ldots s_n$，使得$\prod_{i=1}^{N} P(s_i|s_{i-1})P(w_i|s_i)$概率最大。

针对这一问题，可采用维特比（Viterbi）算法进行求解：

1. **初始化**：对第一个字，分别以BEMS四种状态计算其概率
2. **递归**：对第 i 个字，遍历四种状态，先计算该状态最可能是由前一时刻的哪个状态转换而来的，再乘以该状态下得到观测值（字）的概率，取最大值。
3. **终止**：在第 n 个字时，取得到的最大概率，并得到最后一个字的状态标签。
4. **回溯**：由最优路径的终点向前，找到各个时刻的最优状态，还原全部标签。

维特比算法的一个实现案例：

初始状态概率矩阵

T	F
0.6	0.4

隐含状态转移概率矩阵

	->T	->F
T	0.7	0.3
F	0.4	0.6

观测状态概率矩阵

	->X	->Y	->Z
T	0.5	0.4	0.1
F	0.1	0.3	0.6

问：观测到显式状态序列依次为X、Y、Z，最可能的隐含状态序列是什么？

- 记S是所有可能的状态的集合，$S = \{s_1, s_2\} = \{T, F\}$
- $I = (s_{i_1}, s_{i_2}, s_{i_3})$是长度为3的状态序列，其对应的观测序列为$W = (w_1, w_2, w_3) = (X, Y, Z)$
- 隐含状态转移概率矩阵为$A = [a_{ij}]_{2\times2} = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$，其中，$a_{ij} = P(i_{t+1} = s_j | i_t = s_i)$ $i = 1,2; j = 1,2$
  表示在时刻 t 处于状态s_i的条件下在时刻 $t+1$ 转移到状态s_j的概率。
- 观测状态概率矩阵为$B = [b_j(k)]_{2\times3} = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$，其中，$b_j(k) = P(w_t = v_k | i_t = s_j)$ $k = 1,2,3; j = 1,2$
  表示在时刻 t 处于状态s_j的条件下生成观测v_k的概率。
- π是初始状态概率向量：$\pi = (\pi_1, \pi_2) = (0.6, 0.4)$，其中，$\pi_i = P(i_1 = s_i)$ $i = 1,2$
  表示在时刻 $t=1$ 处于状态s_i的概率

（1）初始化

在$t=1$时，对每一个状态s_i（记显式状态T为s_1，F为s_2），计算状态为s_i观测w_1为X的概率，记此概率为$\delta_1(s_i)$，则

$$\delta_1(s_i) = \pi_i b_i(w_1) \quad i = 1,2$$

代入实际数据

$$\delta_1(T) = 0.6 \times 0.5 = 0.3$$

$$\delta_1(F) = 0.4 \times 0.1 = 0.04$$

可见，第一个隐状态为T更为合理

（2） 在$t=2$时，对每一个状态s_i，求在$t=1$时状态为s_j观测为X并在$t=2$时状态为s_i观测w_2为Y的路径的最大概率，记此最大概率为$\delta_2(s_i)$，则

$$\delta_2(s_i) = \max_{1\le j\le2}\{\delta_1(s_j)a_{ji}\}b_i(w_2)$$

同时，对每个状态s_i，记录概率最大路径的前一个状态s_j所对应的下标 j：

$$\psi_2(s_i) = \arg\max_{1\le j\le2}\{\delta_1(s_j)a_{ji}\} \qquad i = 1,2$$

计算：

$$\delta_2(T) = \max_{1\le j\le2}\{\delta_1(s_j)a_{j1}\}b_1(w_2) = \max_j\{0.3\times0.7, 0.04\times0.4\}\times0.4 = 0.084$$

$$\delta_2(F) = 0.027$$

$$\psi_2(T) = \arg\max_{1\le j\le2}\{\delta_1(s_j)a_{j1}\} = \arg\max_j\{0.3\times0.7, 0.04\times0.4\} = 1$$

$$\psi_2(F) = 1$$

显然，第二个隐状态为T的概率仍然高于F的概率

同理，在$t=3$时，

$$\delta_3(T) = \max_{1\le j\le2}\{\delta_2(s_j)a_{j,1}\}b_1(w_3) = \max_j\{0.084\times0.7, 0.027\times0.4\}\times0.1 = 0.00588$$

$$\psi_3(T) = \arg\max_{1\le j\le2}\{\delta_2(s_j)a_{j,1}\} = \arg\max_j\{0.084\times0.7, 0.027\times0.4\} = 1$$

$$\delta_3(F) = 0.01512$$

此时，第二个隐状态为F的概率更高

$$\psi_3(F) = 1$$

（3）以P^*表示最优路径的概率，则

$$P^* = \max_{1\le i\le2}\delta_3(s_i) = 0.01512$$

最优路径的终点下标i_3^*是 $i_3^* = \arg\max_i\{\delta_3(s_i)\} = 2$

（4）由最终路径的终点下标i_3^*，逆向找到i_1^*, i_2^*：

在t=2时，$i_2^* = \psi_3(s_{i_3^*}) = \psi_3(F) = 1$

在t=1时，$i_1^* = \psi_2(s_{i_2^*}) = \psi_2(T) = 1$

于是求得 $(i_1^*, i_2^*, i_3^*) = (1,1,2)$

最优状态序列 $I^* = (s_1, s_1, s_2) = (T, T, F)$

条件随机场模型（CRF）：具有表达长距离依赖性和交叠性特征的能力，所有特征可以进行全局归一化，能够求得全局的最优解。（隐马尔可夫模型的独立性假设难以描述字词之间的复杂关联）

长短时记忆模型（LSTM）：通过四层神经网络代替 RNN 中原有的单一神经网络层，使其拥有增加或减少信息的能力

通过结合 LSTM 与 CRF 技术，将有效利用句子 级别的标记信息。 • 输出的将不再是相互独立的标签，而是最佳且合理的标签序列。