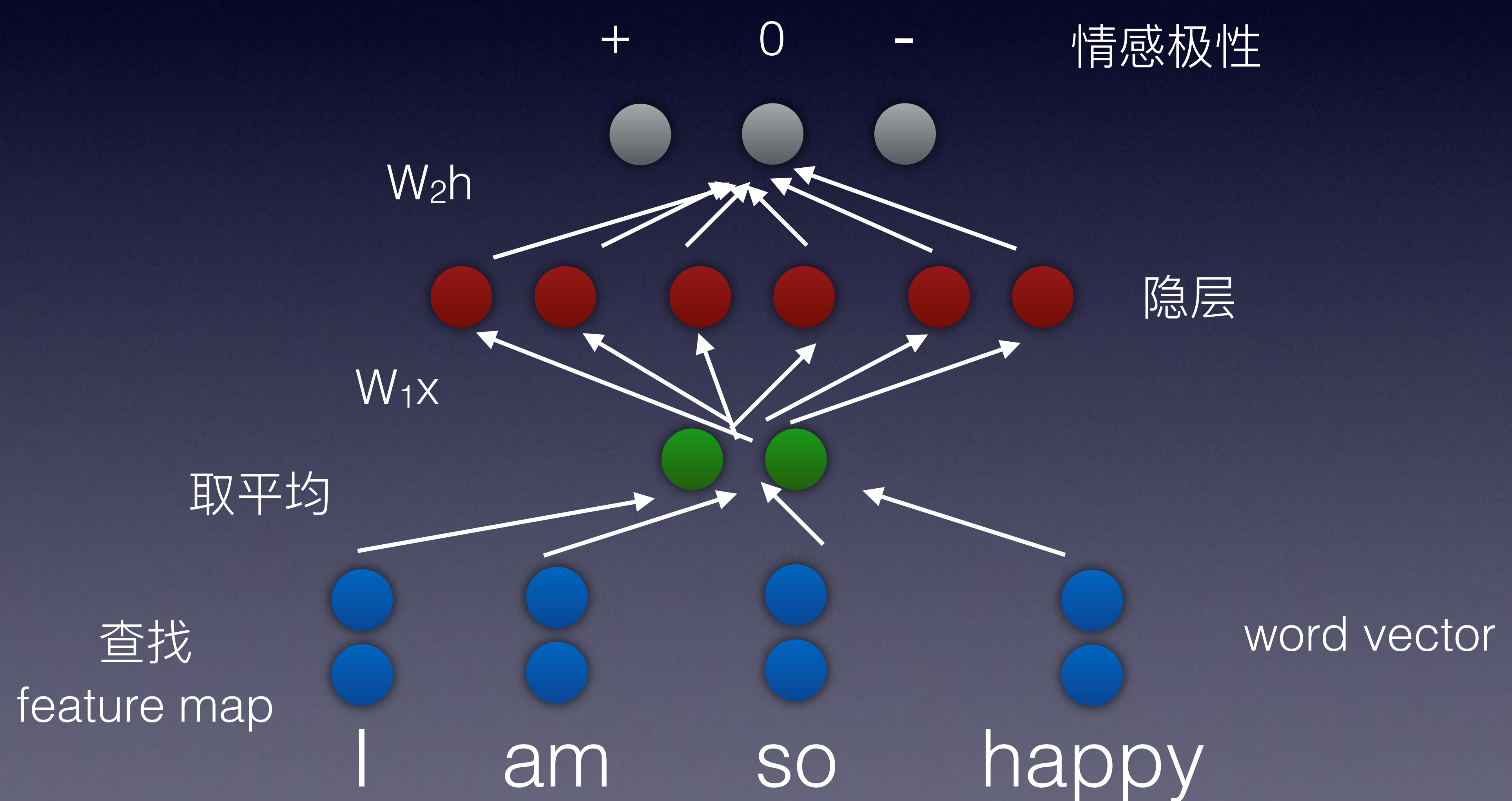
Dynamic Memory Network

Bingfeng Luo

神经网络基本架构

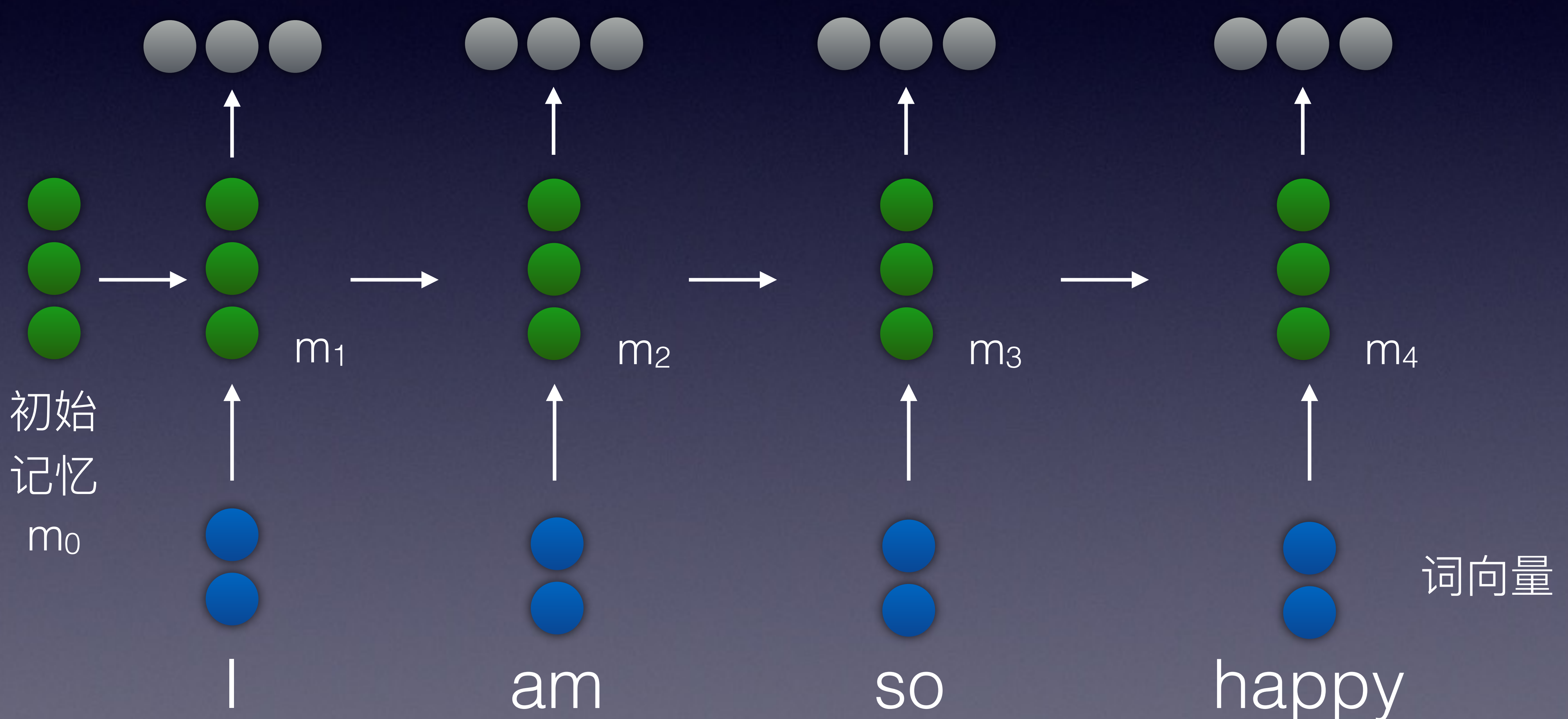
句子情感分析任务



RNN循环神经网络 基本架构

输出，目前为止所能感受到的情感极性

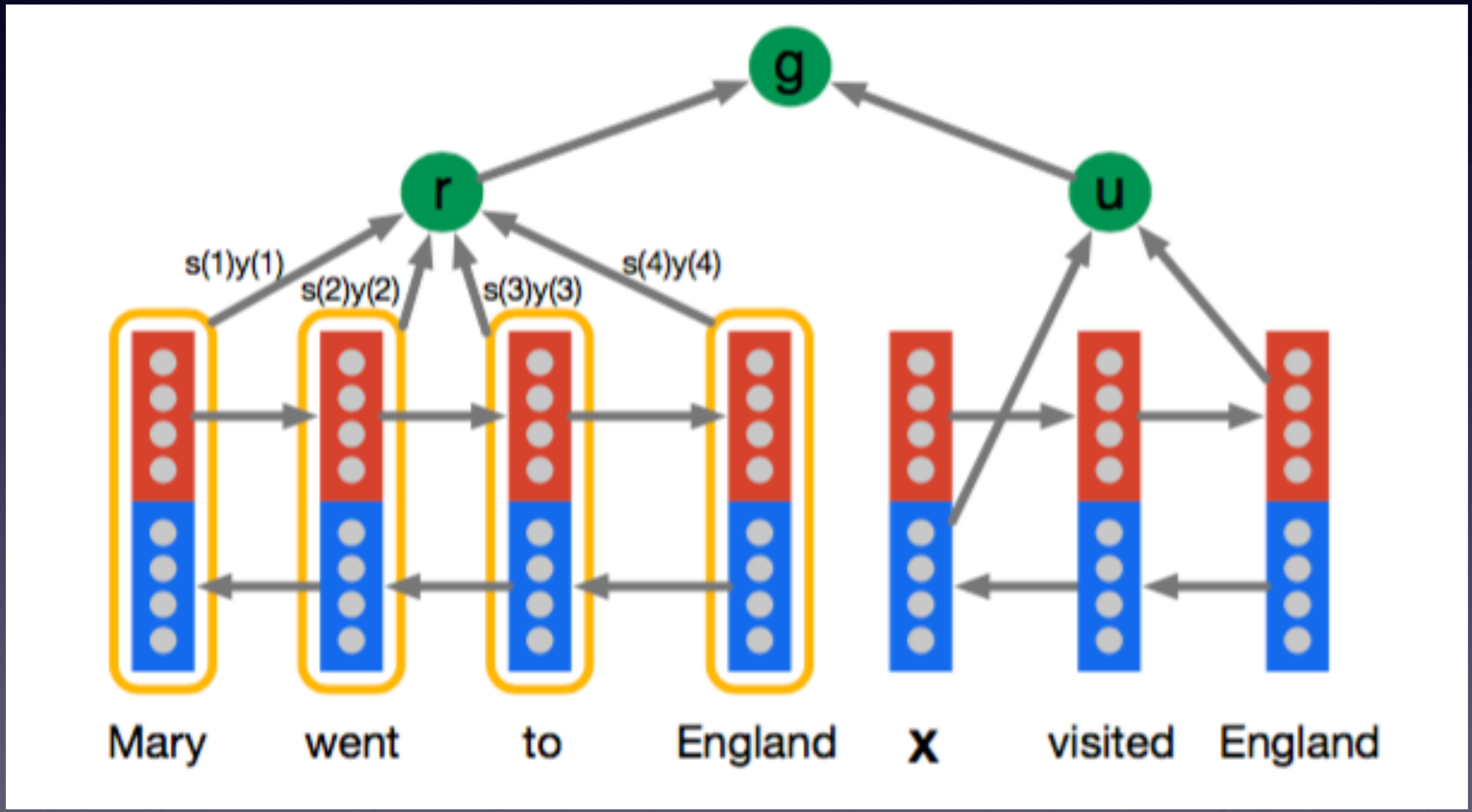
著名的RNN：GRU，LSTM



Attention Network 基本架构

让机器做阅读理解题

- 首先用双向LSMT得到每个词的向量表示
- 然后对每个词，根据query的向量表示和整个document的向量表示，以及当前词的向量表示，计算当前词的权重（attention）
- 最后根据attention加权组合得到r

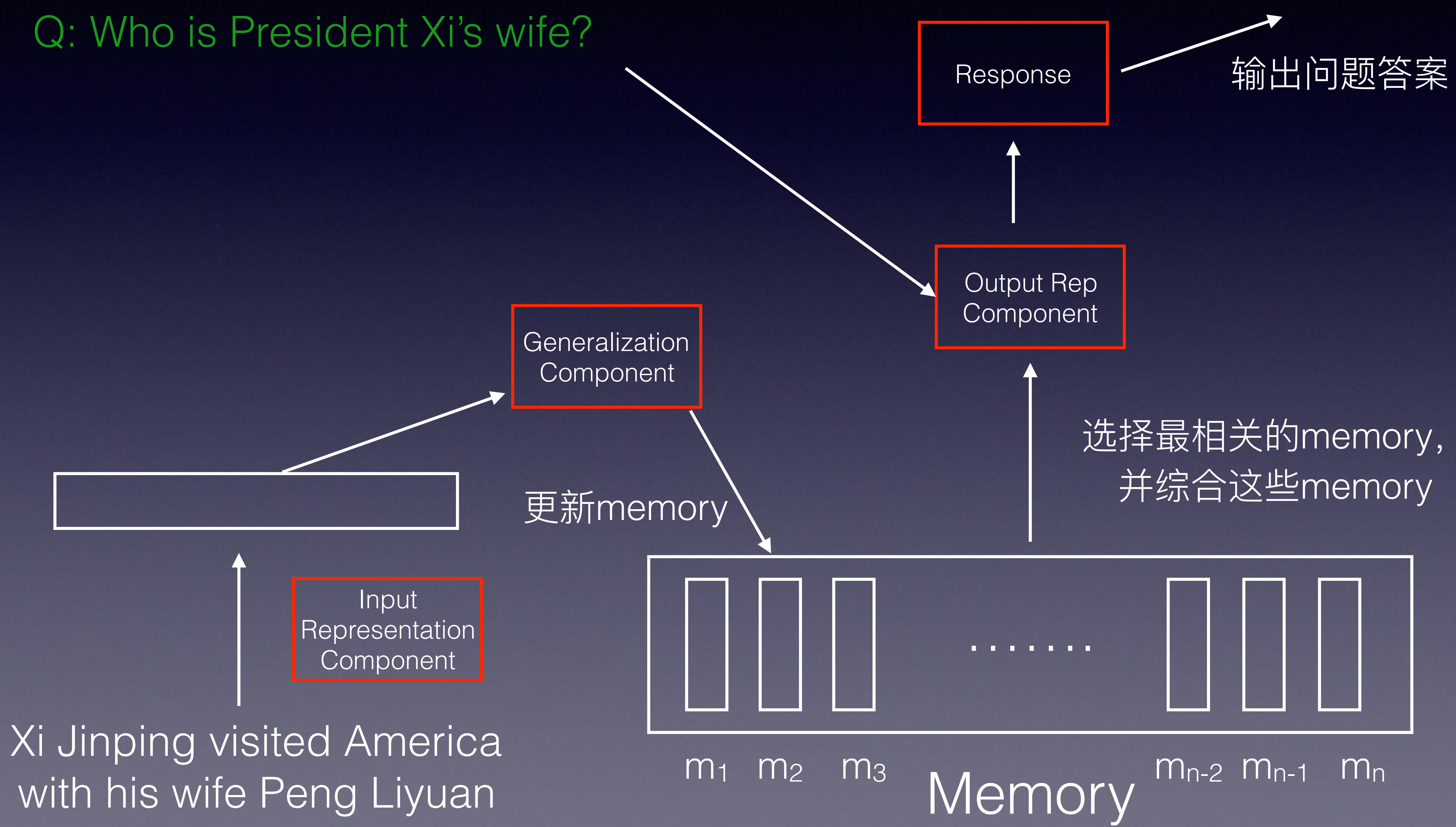


Teaching Machines to Read and Comprehend，DeepMind, 2015

Memory Network 基本架构

- I（Input Feature Map): 将输入转换为representation
- G（Generalization）：根据新的输入，更新memory。这里network很可能会将input压缩和泛化
- O（Output Feature Map）：得到输出的representation
- R（Response）：根据输出的representation得到最终的预测结果

Memory Network 基本架构



Memory Network 例子

MEMORY NETWORKS，Facebook，ICLR 2015

Joe went to the kitchen. Fred went to the kitchen. Joe picked up the milk.
Joe travelled to the office. Joe left the milk. Joe went to the bathroom.
Where is the milk now? A: office
Where is Joe? A: bathroom
Where was Joe before the office? A: kitchen

- 首先扫描所有的facts，得到representation后放到memory的一个slot中
- 找到跟问题的representation最相近的两个memory
- 综合这两个memory，算出output representation
- 根据output representation，和词向量，找到最优的词。如果答案是句子，则用LSTM通过语言模型输出

Memory Network 鼻祖

- Neural Turing Machine，DeepMind，2014
- 读取input和更新memory可以对应图灵机的读写头
- memory可以类比计算机的内存，或者人类的working memory
- 而如果用LSTM来处理memory的话（Output Rep Component），则LSTM的内部记忆可以类比计算机的寄存器
- Neural Turing Machine可以成功的学习copy操作和随机存取操作（indexing），并且可以不是很完美的学会排序操作

bAbI Dataset

- 类似于早期的计算机文字冒险游戏
- 预先在虚拟世界中设定一些entity，并指定这些实体之间的一些action。实体之间可以通过action互动
- 比如Mary takes the football to the kitchen.
- 生成数据集时，首先让这些实体做一些动作，然后根据一些规则问一些问题
- 比如Where is the football
- 通过同义词集来使得自动生成的问题更加符合人类语言
- 设定了一些不同类型的问题，来衡量QA系统多方面的能力，比如推理能力等

bAbI Dataset

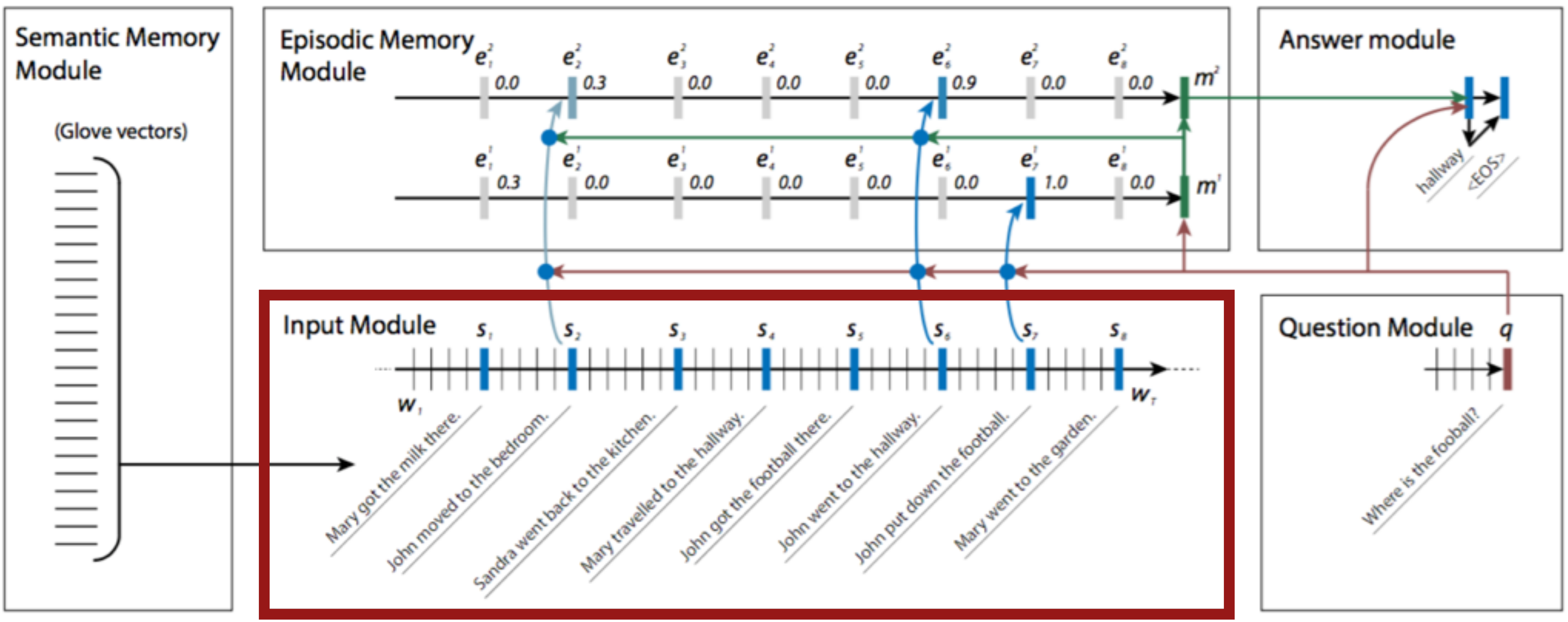
The office is north of the bedroom.
The bedroom is north of the bathroom.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Subject Object Relation

Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

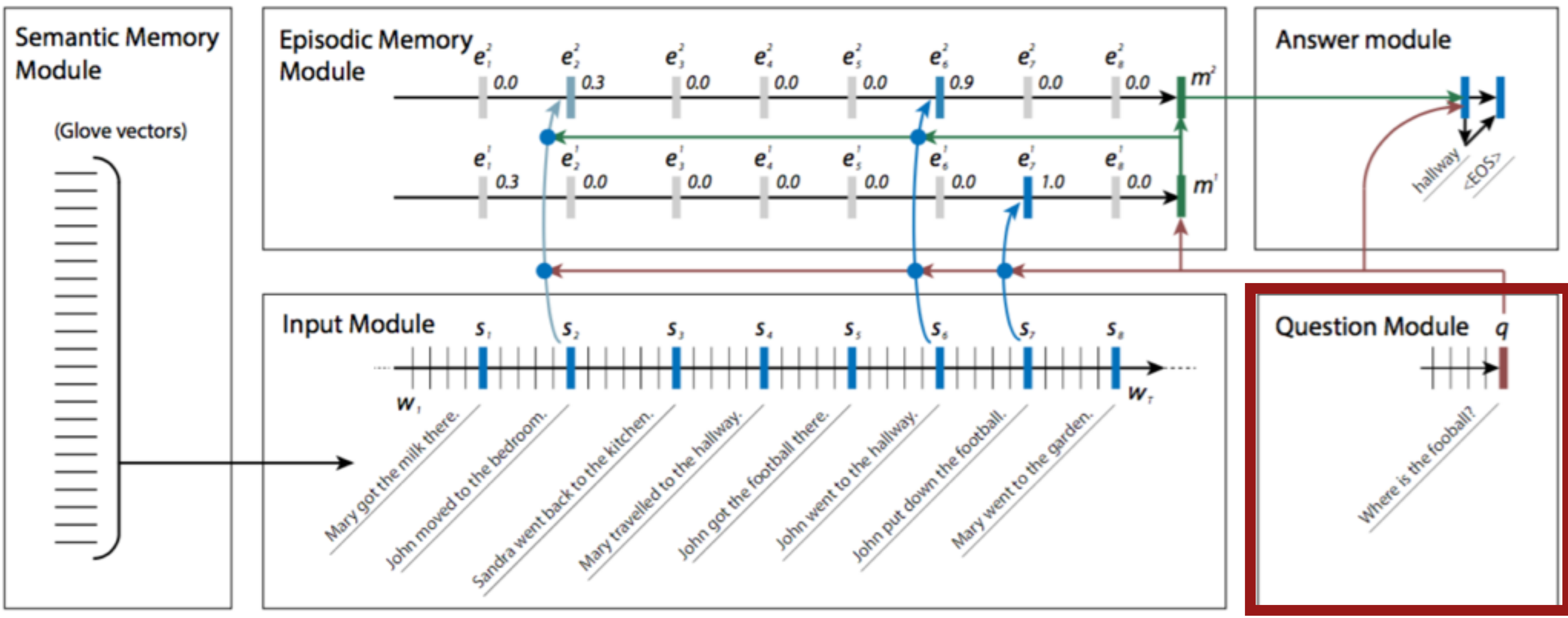
Counting

Dynamic Memory Network



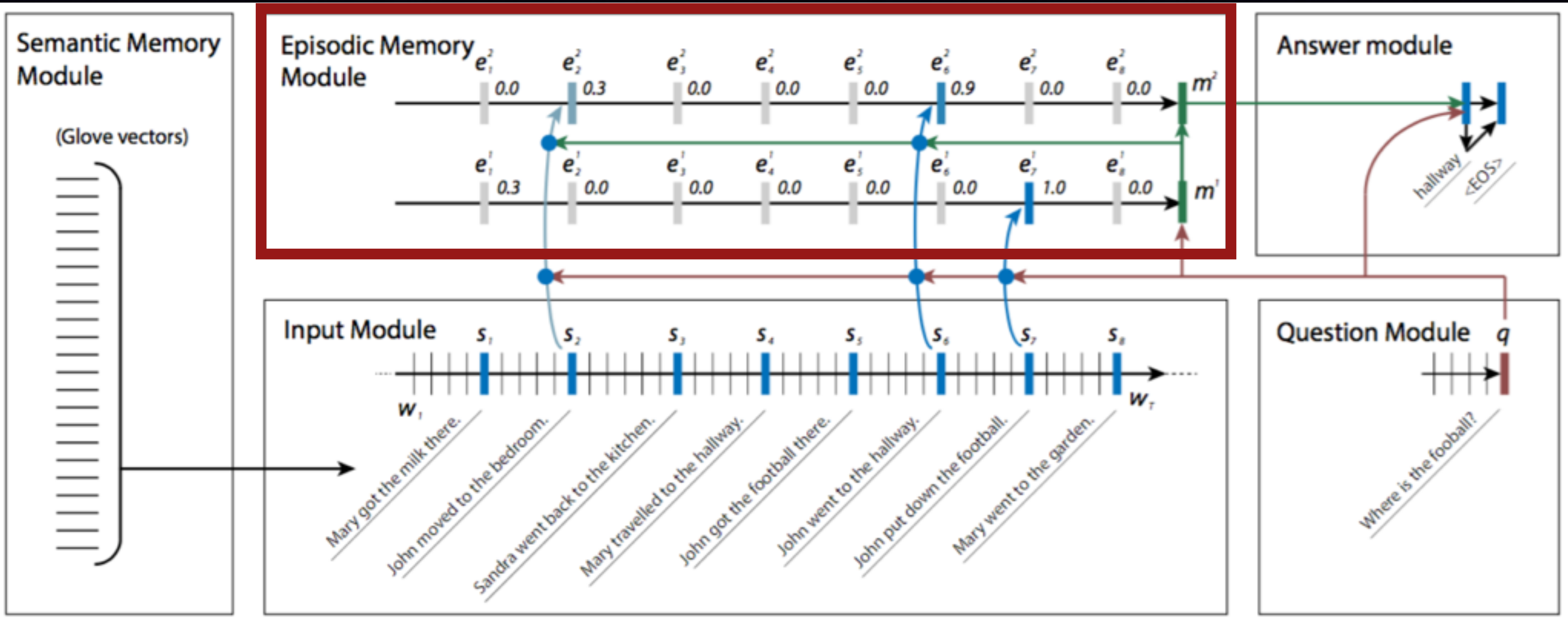
- 输入的句子都通过GRU（RNN的一种）转换成向量，备用
- 可以理解为把每个句子都放进了一个memory slot中

Dynamic Memory Network



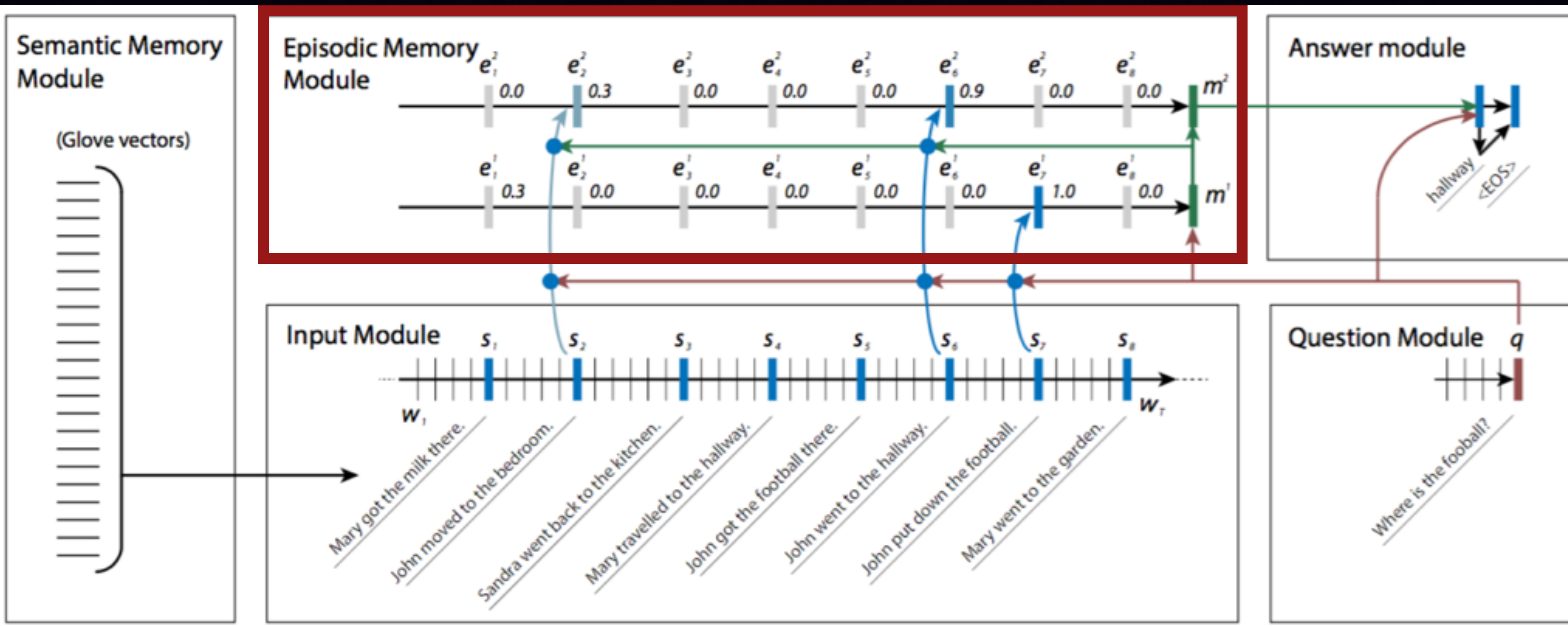
- Question也通过GRU转换成向量，备用

Dynamic Memory Network



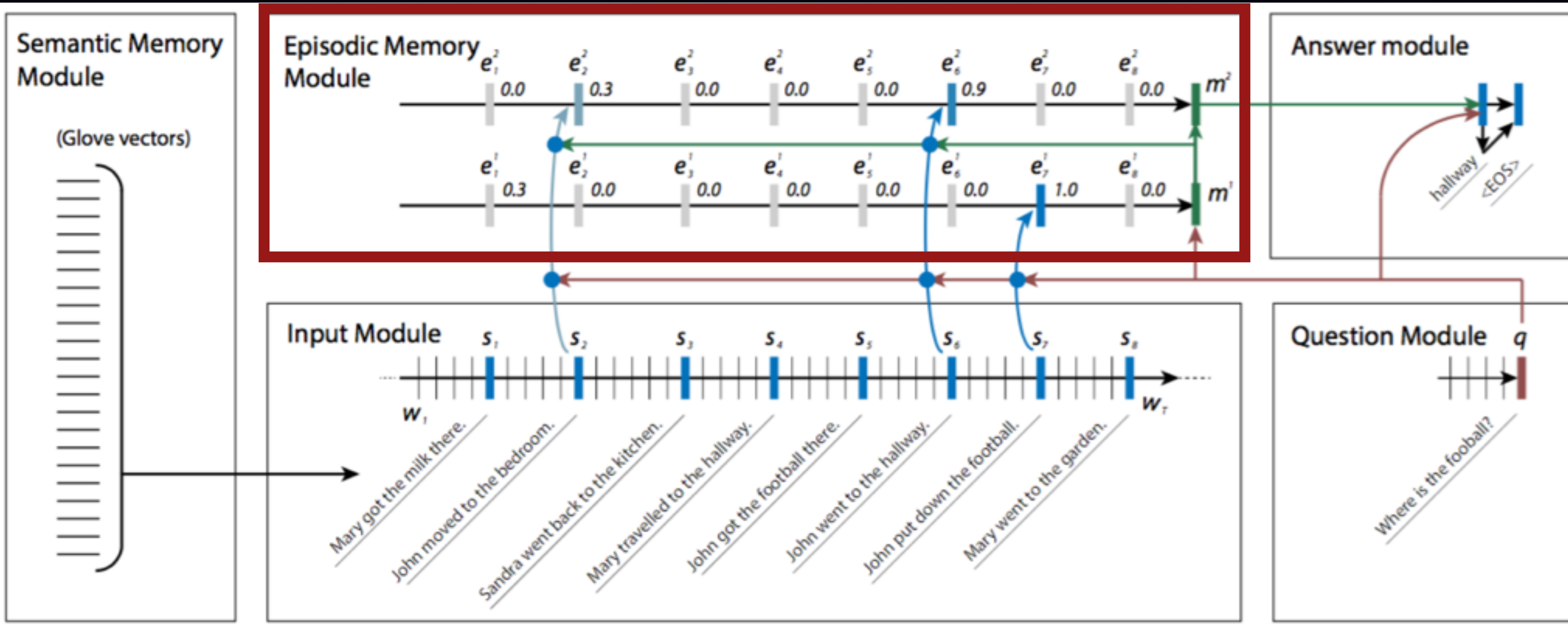
- 引入episodic memory，即对于一个问题，会扫描k次所有的facts，一次扫描产生一个episodic memory

Dynamic Memory Network



- 扫描时，使用GRU进行扫描（相当于一个序列），对每一个fact算一个gate值（attention），并用这个gate值来辅助更新GRU的memory

Dynamic Memory Network



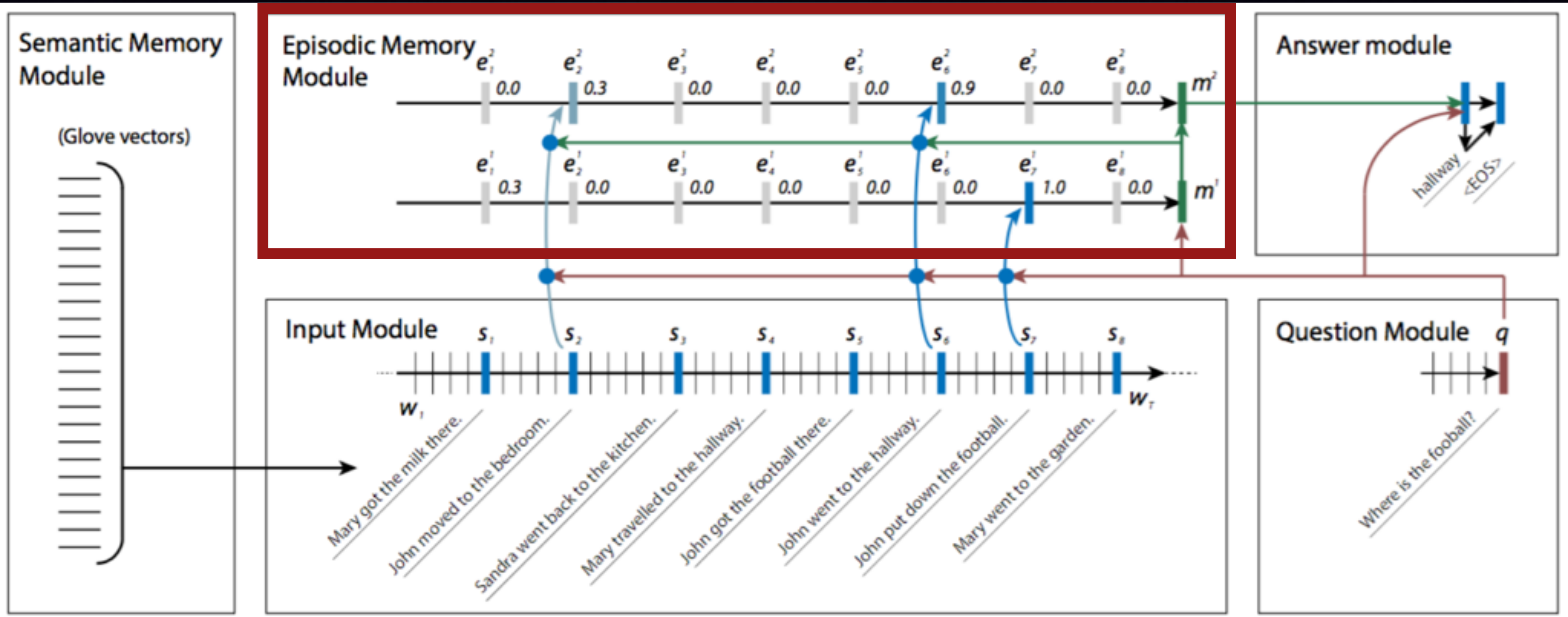
- 设总共扫描了k次，则我们可以对这个长度k的序列再用GRU扫描一遍，得到最终memory，作为answer模块的输入

Dynamic Memory Network

- 设m^i是用来扫描episode memory的GRU的内部memory，即其包含了之前i次扫描的所有记忆
- 我们用m^{i-1}，当前fact的向量c，question的向量q来计算当前fact的gate（目前看来是否有用，attention）
- 具体计算方法为：

$$
\begin{aligned}
z(c,m,q) &= [c, m, q, c \circ q, c \circ m, |c-q|, |c-m|, c^T W^{(b)} q, c^T W^{(b)} m] \\
G(c,m,q) &= \sigma\left(W^{(2)} \tanh\left(W^{(1)} z(c,m,q) + b^{(1)}\right) + b^{(2)}\right)
\end{aligned}
$$

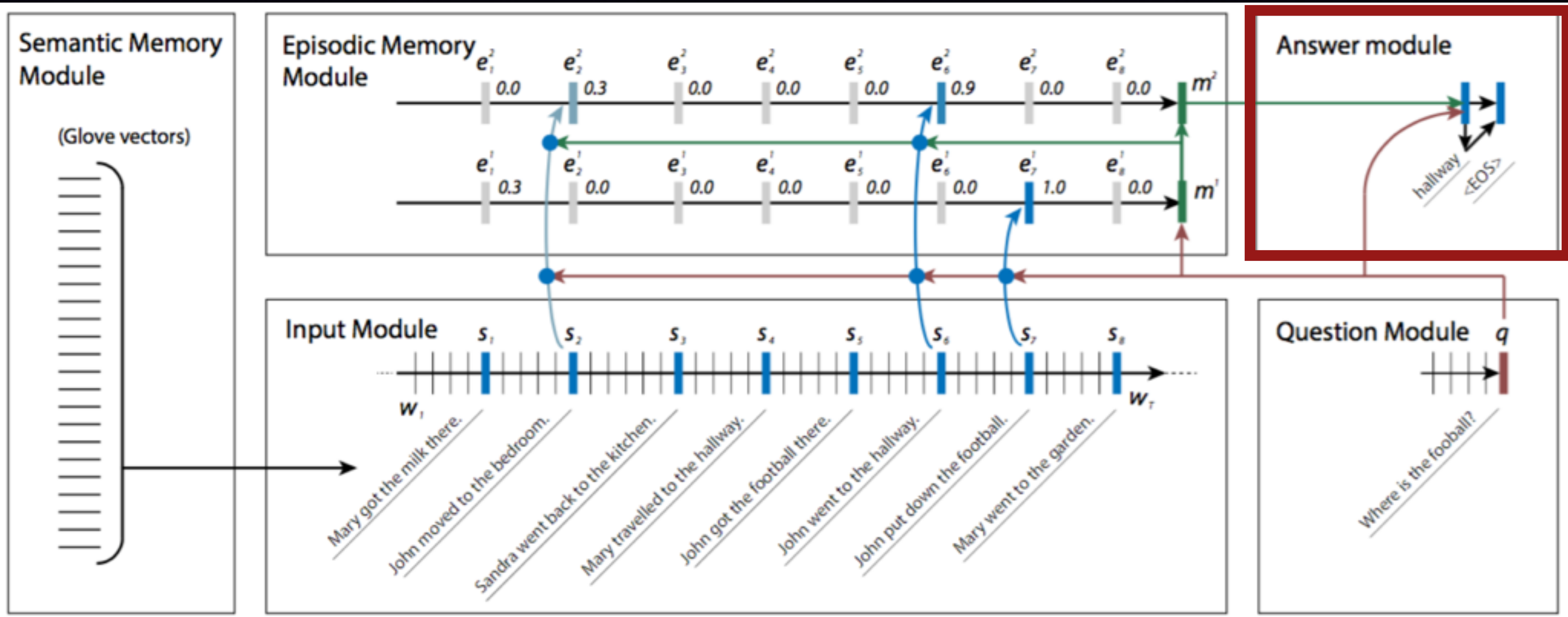
Dynamic Memory Network



- 之前的gate被用来指导扫描facts的GRU的更新：

$$h_t^i = g_t^i GRU(c_t, h_{t-1}^i) + (1 - g_t^i) h_{t-1}^i$$

Dynamic Memory Network



- answer模块就是再用一个GRU来decode之前生成的memory向量即可

Dynamic Memory Network

- 训练：
- 虽然整个模型复杂无比，但是可以发现每个部分都是可微的，所以可以用梯度下降来训练
- 由于模型极其复杂，所以必须用模块化的方式实现

Facebook bAbI Dataset

Task	MemNN	DMN	Task	MemNN	DMN
1: Single Supporting Fact	100	100	11: Basic Coreference	100	99.9
2: Two Supporting Facts	100	98.2	12: Conjunction	100	100
3: Three Supporting facts	100	95.2	13: Compound Coreference	100	99.8
4: Two Argument Relations	100	100	14: Time Reasoning	99	100
5: Three Argument Relations	98	99.3	15: Basic Deduction	100	100
6: Yes/No Questions	100	100	16: Basic Induction	100	99.4
7: Counting	85	96.9	17: Positional Reasoning	65	59.6
8: Lists/Sets	91	96.5	18: Size Reasoning	95	95.3
9: Simple Negation	100	100	19: Path Finding	36	34.5
10: Indefinite Knowledge	98	97.5	20: Agent's Motivations	100	100
			Mean Accuracy (%)	93.3	**93.6**

POS Tagging

Model	SVMTool	Sogaard	Suzuki et al.	Spoustova et al.	SCNN	DMN
Acc (%)	97.15	97.27	97.40	97.44	97.50	**97.56**

情感分析

Task	MV-RNN	RNTN	DCNN	PVec	CNN-MC	DRNN	CT-LSTM	DMN
Binary	82.9	85.4	86.8	87.8	88.1	86.6	88.0	**88.3**
Fine-grained	44.4	45.7	48.5	48.7	47.4	49.8	**51.0**	50.3

早期版本的结果

Task	MV-RNN	RNTN	DCNN	PVec	CNN-MC	DRNN	CT-LSTM	DMN
Binary	82.9	85.4	86.8	87.8	88.1	86.6	88.0	**88.6**
Fine-grained	44.4	45.7	48.5	48.7	47.4	49.8	51.0	**51.2**

九月底的结果

指代消解

Metric	Guha et al., 2015			Durrett and Klein, 2013			DMN		
	P	R	F1	P	R	F1	P	R	F1
MUC	56.8	57.8	57.8	70.2	40.2	49.6	74.6	66.0	**70.0**
B^3	68.1	74.8	70.4	88.5	64.7	74.2	82.0	77.5	**79.7**
$CEAF_e$	73.3	76.1	74.2	56.5	80.0	65.7	70.7	79.3	**74.7**

Q&A