浅谈人类记忆和计算机存储系统在信息提取方面的异同

心理学求是科学班2201 李烁安

## 

摘 要 人类记忆和计算机存储系统的功能都是编码、存储和提取信息。二者在结构 和机制等层面上有很多共同点，也有相应的差异性。本文着重于信息提取方 面对二者进行比较，从提取的阶段、信息表征方法、检索机制、提取的影响 因素等方面讨论二者的相似点和差别。

关键词 记忆，计算机存储，信息提取，信息表征，语义启动模型

## 1 背景

### 1.1 共同的问题

人类记忆和计算机存储系统的基本功能都是编码、存储和提取信息。人类和计算机的很多活动和操作是基于对已编码、存储的信息的提取（retrieval）的，如果信息提取不够快速，人类和计算机都将无法对刺激做出及时的反应。另一方面，如果提取的信息是错误信息，将会导致错误的行为，或导致需要更多额外的时间去找到正确的信息。所以，人类记忆和计算机信息系统共有的一个问题就是如何快速而精确地提取信息。

### 1.2 比较二者的目的

理论心理学对于记忆提取过程的研究主要着眼于了解信息是如何在人类记忆系统中被表征和检索的。计算机科学家们则希望优化信息的编码方式和检索算法来构建更好的计算机检索系统。

研究人类记忆系统提取信息的方法有助于我们从仿生的角度优化计算机信息检索算法。而计算机检索系统的发展可以启发我们对与人类记忆系统的模型构建。

对两者进行比较可以更好地认识两者本身的结构、机制等，并最终起到取长补短的效果，助推心理学和计算机科学的共同发展。

### 1.3 声明

本文讨论的信息提取指的是基于一定线索（cue）的有目的的提取过程，且主要针对陈述性记忆而非程序性记忆。本文讨论的计算机存储涉及到人类的主观意图，比如利用搜索引擎进行信息检索，不仅限于单个计算机内部的存储。

## 2 两者的共同点

### 2.1 类似的阶段

人类记忆和计算机存储系统在提取信息时都涉及到三个阶段[1]：

1、获取（generate）和汇总（assemble）线索；

2、依靠线索检索信息；

3、验证（verify）信息是否是想要的信息。

已知的大多数人类记忆系统信息提取的模型都是基于线索激活的，识别线索、整理线索是记忆提取的开端。计算机中，对信息的提取也是从人为或自发的线索开始的，这种线索可以是文件地址、正则表达式、关键词等。

在获取和汇总线索的基础上，两者对希望得到的信息进行检索，检索的过程涉及到不同的模型和算法，是信息提取过程最核心也是最复杂的一环。

验证则是为保证提取信息的正确性必不可少的一环。无论是人类记忆还是计算机存储，验证主要是基于对已检索到的信息和线索的比较，这种比较可以是绝对的字面的比较，也可以是基于特征加权的可能性比较。

事实上，信息提取的过程是包含这三个阶段的条件循环。获得线索后进行检索，验证检索到的信息，若通过则停止提取过程，若不通过则继续整理线索，再次检索、验证。

### 2.2 类似的信息表征方式

网络（network）和特征（feature）是人类记忆和计算机存储系统中信息表征的通用手段[2]。

#### 2.2.1 网络表征

在网络表征中，信息以离散的概念（concept）形式被存储在记忆节点中，这些节点以网络的形成互相关联。网络表征相比线性表征的优势在于，节点数量庞大且节点间关系复杂的复杂网络（complex network）系统会体现涌现（emergence）特征和小世界现象（"small world" phenomenon），大大增加了节点之间的通达度[3]。

这种表征方式在语义记忆的激活扩散模型（spreading activation model）[4]中得到了体现。基于激活扩散模型，也诞生了针对计算机存储系统的信息检索模型[5]。比如搜索引擎基于超链接的检索，就利用了不同网站间的双向链接作为关联度计算的重要依据。

另一种基于网络表征的信息提取模型是复合线索模型（compound cue model）[6]。该模型认为已激活的节点会在工作记忆中和已有的线索复合成为新的线索，进一步激活更多的节点。这种激活方式更加线性，更强调深度而非广度。搜索引擎基于关键词的全文检索技术就利用了类似的模型来加强检索结果的精确性。计算机硬盘存储的提取过程也强调类似的“长时记忆”与“工作记忆”结合的提取模式，即通过RAM暂存已提取的信息和线索以提高检索效率，减少检索过程对存储本身的干扰。

#### 2.2.2 特征表征

记忆的特征表征以分布记忆模型（distributed model）[7]为代表。这种表征方式将信息视为描述其特征的信息组成的向量，信息通过类似的特征被激活和筛选。这种表征方式同样常见于计算机中的数据库系统，数据库中的每一个记录由不同字段组成，字段相当于记录的特征，被用于筛选式的检索。这样的模式也可以抽象为信息在计算机中存储时所附带的各种元信息（metadata），如存储时间、类型、创建者等。

## 3 两者的差异

### 3.1 建立关联的能力不同

基于网络的表征方式依靠信息的关联度提高检索信息的效率。人类和计算机建立信息关联的能力是很不同的[1-1]。

人类可以自发地依靠语义和概念的抽象特征建立信息节点的关联。比如“土豆”和“马铃薯”会被视为同样的概念，“苹果”和“香蕉”会自然地通过“水果”这一上层概念建立起关联。相比之下，计算机不能自发建立信息之间的抽象关联，且针对语义的关联主要依靠重复的字词而非语义来建立。

### 3.2 利用情景信息的能力不同

人类的陈述性记忆包括情景记忆，这种记忆方式可以高效地利用情境信息（环境、情绪等）来记忆、检索信息[8]。而计算机只能依赖最基础的情景信息，比如信息记录时间、创建者等。

### 3.3 消退的能力不同

如果不经过规律的复述、再认，人类记忆是会随着时间推移而逐渐消退的。而计算机存储的信息不会自发消退，只要硬盘寿命允许就可以永久保存。

记忆消退实际上不绝对是坏事[9]。因为如果记忆难以消退，新的信息就会触发过多已记忆的内容，使形成新记忆的过程容易受到干扰。而且过多不必要的信息激活也会影响检索的效率和精确度。事实上，为了避免这样的影响，计算机信息检索系统常通过针对存储时间加权来减少老旧信息被激活的概率。

## 参考文献

1. Foltz, P.W. (1991). Models of Human Memory and Computer Information Retrieval: Similar Approaches to Similar Problems.
2. Liuzzi, A.G., Aglinskas, A. & Fairhall, S.L.(2020). General and feature-based semantic representations in the semantic network.
3. Mitchell, Melanie (2009). Complexity: A Guided Tour.
4. Collins, Allan M.; Loftus, Elizabeth F. (1975). "A spreading-activation theory of semantic processing".
5. Crestani, F. (1997). Application of Spreading Activation Techniques in Information Retrieval.
6. Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory.
7. Masson, M. E. J. (1995). A distributed memory model of semantic priming.
8. Pessin, J. (1932). The effect of similar and dissimilar conditions upon learning and relearning.
9. Gilbert, L. T., Delaney, P. F., & Racsmány, M. (2022). People sometimes remember to forget: Strategic retrieval from the list before last enables directed forgetting of the most recent information.