【论文导读】大语言模型综述（一）：介绍

@Sa神带你学AI

Info

视频简介
本系列为《A Survey of Large Language Model》的论文导读系列视频，本视频导读内容为论文的第一章，即Introduction部分。
本次介绍内容主要包括语言模型发展的四个阶段、大语言模型的突出能力、大语言模型的进展和挑战。
论文引用：W. X. Zhao et al., 'A Survey of Large Language Models'. arXiv, Nov. 24, 2023. doi: 10.48550/arXiv.2303.18223.
推荐阅读：
<https://github.com/RUCAIBox/LLMSurvey/blob/main/assets/LLM_Survey_Chinese.pdf>
笔记注记：
黄色：名词组分
粉色：动词组分
绿色：状语和修饰语组分
橙色：专有名词

1 Introduction

"The limits of my language mean the limits of my world."
—Ludwig Wittgenstein

Language is a prominent ability in human beings to express and communicate, which develops in early childhood and evolves over a lifetime [3, 4]. Machines, however, cannot naturally grasp the abilities of understanding and communicating in the form of human language, unless equipped with powerful artificial intelligence (AI) algorithms. It has been a longstanding research challenge to achieve this goal, to enable machines to read, write, and communicate like humans [5].

Technically, *language modeling (LM)* is one of the major approaches to advancing language intelligence of machines. In general, LM aims to model the generative likelihood of word sequences, so as to predict the probabilities of future (or missing) tokens. The research of LM has received extensive attention in the literature, which can be divided into four major development stages:

- *Statistical language models (SLM).* SLMs [6–9] are developed based on *statistical learning* methods that rose in the 1990s. The basic idea is to build the word prediction model based on the Markov assumption, *e.g.*, predicting the next word based on the most recent context. The SLMs with a fixed context length n are also called n-gram language models, *e.g.*, bigram and trigram language models. SLMs have been widely applied to enhance task performance in information retrieval (IR) [10, 11] and natural language processing (NLP) [12–14]. However, they often suffer from the curse of dimensionality: it is difficult to accurately estimate high-order language models since an exponential number of transition probabilities need to be estimated. Thus, specially designed smoothing strategies such as back-off estimation [15] and Good–Turing estimation [16] have been introduced to alleviate the data sparsity problem.

- *Version: v13 (major update on November 23, 2023).*

【带注释】A Survey of Large Language Model.pdf

Further Reading

"The limits of my language mean the limits of my world."
—Ludwig Wittgenstein

路德维希·约瑟夫·约翰·维特根斯坦(1889-1951)是一位奥地利哲学家，主要研究逻辑学、数学哲学、心灵哲学和语言哲学。[1]

1. WikiPedia - Ludwig Wittgenstein