

# 机器翻译

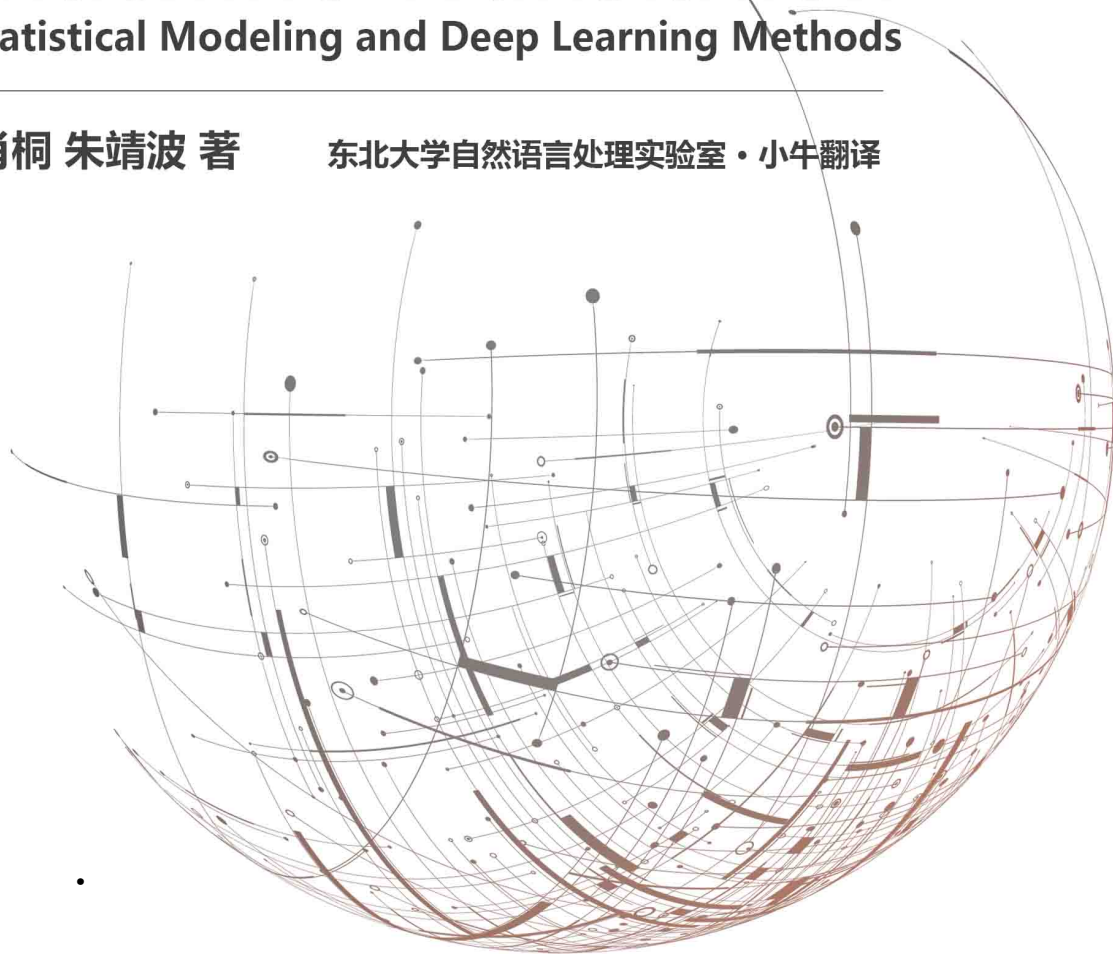
统计建模与深度学习方法

## Machine Translation

Statistical Modeling and Deep Learning Methods

肖桐 朱靖波 著

东北大学自然语言处理实验室 · 小牛翻译



Copyright © 2020 肖桐 朱靖波

东北大学自然语言处理实验室 · 小牛翻译

[HTTPS://OPENSOURCE.NIUTRANS.COM/MTBOOK/INDEX.HTML](https://opensource.niutrans.com/mtbook/index.html)

[HTTPS://GITHUB.COM/NIUTRANS/MTBOOK](https://github.com/NiUTrans/MTBook)

Licensed under the Creative Commons Attribution-NonCommercial 4.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/4.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “as is” basis, without warranties or conditions of any kind, either express or implied. See the License for the specific language governing permissions and limitations under the License.

*May 8, 2020*

在此感谢所有为本书做出贡献的人

曹润柘、曾信、孟霞、单韦乔、姜雨帆、王子扬、刘辉、许诺、李北、刘继强、张哲旻、周书含、周涛、张裕浩、李炎洋、林野、刘晓倩、牛蕊、田丰宁、杜权

# 导 读

让计算机进行自然语言的翻译是人类长期的梦想，也是人工智能的终极目标之一。自上世纪九十年代起，机器翻译迈入了基于统计建模的时代，发展到今天，深度学习等机器学习方法已经在机器翻译中得到了大量的应用，取得了令人瞩目的进步。

在这个时代背景下，对机器翻译的模型、方法和实现技术进行深入了解是自然语言处理领域研究者和实践者所渴望的。本书全面回顾了近三十年内机器翻译的技术发展历程，并围绕统计建模和深度学习两个主题对机器翻译的技术方法进行了全面介绍。在写作中，笔者力求用朴实的语言和简洁的实例阐述机器翻译的基本模型和方法，同时对相关的技术前沿进行讨论。本书可以供计算机相关专业高年级本科生及研究生学习之用，也可以作为自然语言处理，特别是机器翻译领域相关研究人员的参考资料。

本书共分为七个章节，章节的顺序参考了机器翻译技术发展的时间脉络，同时兼顾了机器翻译知识体系的内在逻辑。各章节的主要内容包括：

- 第一章：机器翻译简介
- 第二章：词法、语法及统计建模基础
- 第三章：基于词的机器翻译模型
- 第四章：基于短语和句法的机器翻译模型
- 第五章：人工神经网络和神经语言建模
- 第六章：神经机器翻译模型
- 第七章：神经机器翻译实战 —— 参加一次比赛

其中，第一章是对机器翻译的整体介绍。第二章和第五章是对统计建模和深度学习方法的介绍，分别建立了两个机器翻译范式的基础知识体系——统计机器翻译和神经机器翻译。统计机器翻译部分（第三、四章）涉及早期的基于单词的翻译模型，以及本世纪初流行的基于短语和句法的翻译模型。神经机器翻译（第六、七章）代表了当今机器翻译的前沿，内容主要涉及了基于端到端表示学习的机器翻译建模方法。特别的，第七章对一些最新的神经机器翻译方法进行了讨论，为相关科学问题的研究和实用系统的开发提供了可落地的思路。图1展示了本书各个章节及核心概念之间的关系。

用最简单的方式阐述机器翻译的基本思想是笔者所期望达到的目标。但是，书中不可避免会使用一些形式化定义和算法的抽象描述，因此，笔者尽所能通过图例进行解释（本书共 320 张插图）。不过，本书所包含的内容较为广泛，难免会有疏漏，望读者海涵，并指出不当之处。

