清华大学学位论文 LaTeX 模板
使用示例文档 v7.1.0

(申请清华大学工学硕士学位论文)

培 养 单 位 ： 计算机科学与技术系

学　　科 ： 计算机科学与技术

研 究 生 ： 薛 瑞 尼

指 导 教 师 ： 郑 纬 民　教 授

副指导教师 ： 陈 文 光　教 授

二〇二一年三月

An Introduction to LaTeX Thesis Template of Tsinghua University v7.1.0

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Master of Science

in

Computer Science and Technology

by

Xue Ruini

Thesis Supervisor: Professor Zheng Weimin

Associate Supervisor: Professor Chen Wenguang

March, 2021

学位论文指导小组、公开评阅人和答辩委员会名单

指导小组名单

李 XX	教授	清华大学
王 XX	副教授	清华大学
张 XX	助理教授	清华大学

公开评阅人名单

刘 XX	教授	清华大学
陈 XX	副教授	XXXX 大学
杨 XX	研究员	中国 XXXX 科学院 XXXXXXX 研究所

答辩委员会名单

主席	赵 XX	教授	清华大学
委员	刘 XX	教授	清华大学
	杨 XX	研究员	中国 XXXX 科学院 XXXXXXX 研究所
	黄 XX	教授	XXXX 大学
	周 XX	副教授	XXXX 大学
秘书	吴 XX	助理研究员	清华大学

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容；（3）按照上级教育主管部门督导、抽查等要求，报送相应的学位论文。

本人保证遵守上述规定。

（保密的论文在解密后遵守此规定）

作者签名：_____　　导师签名：_____

日　　期：_____　　日　　期：_____

摘　要

论文的摘要是对论文研究内容和成果的高度概括。摘要应对论文所研究的问题及其研究目的进行描述，对研究方法和过程进行简单介绍，对研究成果和所得结论进行概括。摘要应具有独立性和自明性，其内容应包含与论文全文同等量的主要信息。使读者即使不阅读全文，通过摘要就能了解论文的总体内容和主要成果。

论文摘要的书写应力求精确、简明。切忌写成对论文书写内容进行提要的形式，尤其要避免“第 1 章……；第 2 章……；……”这种或类似的陈述方式。

关键词是为了文献标引工作、用以表示全文主要内容信息的单词或术语。关键词不超过 5 个，每个关键词中间用分号分隔。

关键词：关键词 1；关键词 2；关键词 3；关键词 4；关键词 5

Abstract

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summarization of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Keywords are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 keywords, with semi-colons used in between to separate one another.

Keywords: keyword 1; keyword 2; keyword 3; keyword 4; keyword 5

目 录

插图和附表清单

符号和缩略语说明

PI	聚酰亚胺
MPI	聚酰亚胺模型化合物，N-苯基邻苯酰亚胺
PBI	聚苯并咪唑
MPBI	聚苯并咪唑模型化合物，N-苯基苯并咪唑
PY	聚吡咙
PMDA-BDA	均苯四酸二酐与联苯四胺合成的聚吡咙薄膜

第 1 章　引言

1.1　研究背景

网络自诞生以来，随着数十年的发展，已经成为了最重要的信息化基础设施之一。如图 1.1所示，第 46 次《中国互联网络发展状况统计报告》[1] 指出，截至 2020 年 6 月，我国网民规模达到 9.40 亿，互联网普及率达 67.0%，互联网应用涵盖即时通讯、搜索引擎、网络新闻、在线教育、购物出行等方面，可以说互联网已经和每个人的生活息息相关密不可分。

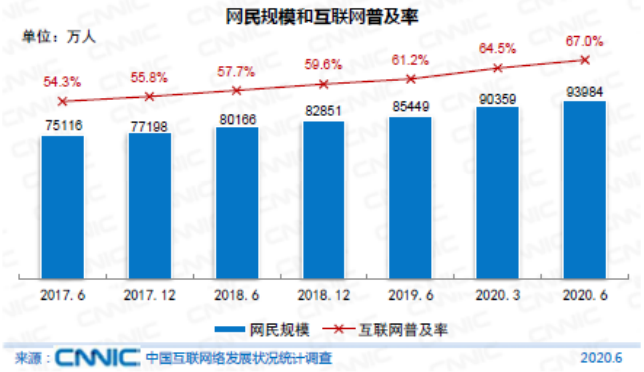

图 1.1　网民规模和互联网普及率

随着网络重要性的提升、用户规模的膨胀，管理网络的难度也越来越大。网络是一个复杂的系统，它的部署、运行和维护都需要专业的运维人员。早期的运维工作大部分是由运维人员手工完成，然后人们逐渐发现一些重复性的工作可以用自动化脚本来实现，于是诞生了自动化运维。自动化运维可以认为是基于专家经验、人为制定规则的系统。但是随着互联网规模急剧膨胀，以及服务类型的多样化，简单的、基于人为制定规则的方法并不能解决大规模运维的问题，因此产生了智能运维。与自动化运维依赖专家知识、人工生成规则不同，智能运维强调使用机器学习算法从海量运维数据中不断学习、不断提炼规则。

异常检测是智能运维的关键环节，具有至关重要的意义。从网络故障管理的角度来说，做好异常检测可以提前预测故障的发生；从性能管理的角度来说，可以发现性能不佳的区域，避免因误配置、架构不合理导致性能下降；从安全管理的角度来说，在网络攻击的前期阶段，及时发现并预警后续攻击，进而做出防御措施。因此，在复杂的网络环境中甄别出有效和异常流量尤为重要，在重大事故发生前，根据各项流量特征的变化，提前预测出即将发生的事故，提高应急响应速度，防患于未然。

1.2　研究内容

1.3　主要贡献

1.4　论文组织结构

第 2 章　相关工作综述

2.1　引言

本章对网络流量异常检测领域的相关工作进行综述。首先介绍了网络流量的特点，

异常检测是一个重要的领域，自 1980 年以来，国内外已经有无数学者在这方面做研究。分类、统计、信息理论和聚类。

Ahmed et al.[2] 将异常检测技术分为分类、统计、信息理论和聚类四类。

本章还讨论了用于网络入侵检测的数据集的研究挑战。

2.2　网络异常的定义和分类

Hawkins(1980) 给出了异常的本质性的定义[3]：异常是在数据集中与众不同的数据，使人怀疑这些数据并非随机偏差，而是产生于完全不同的机制。例如在某个季节里，某一天的气温很高或很低，这个温度数据就是一个异常。网络异常是指那些可能改变网络流量特征或统计指标的恶意行为。网络异常大致可以分为物理故障，网络扫描，BGP 前缀劫持，拒绝服务攻击，蠕虫攻击等几种类型。

2.3　异常检测

目前，学术界和工业界已经提出了一系列 KPI 异常检测算法。这些算法可以概括地分成基于窗口的异常检测算法，例如奇异谱变换 (singular spectrum transform)；基于近似性的异常检测算法；基于预测的异常检测算法，例如 Holt-Winters 方法、时序分解方法、线性回归方法、支持向量回归等；基于隐式马尔科夫模型的异常检测算法；基于分段的异常检测算法；基于机器学（集成学习的异常检测算法等类别。

2.3.1　基于时间序列的异常检测

时间序列是将某种统计指标的数值，按时间先后顺序排序所形成的数列。时间序列的预测就是通过分析时间序列，根据时间序列所反映出来的发展过程、方向和趋势，进行类推或延伸，预测下一段时间或以后若干年内可能达到的水平。时间序列的异常检测就是通过历史的数据分析，查看当前的数据是否发生了明显偏离了正常的情况。主要的时间序列模型有移动平均、指数平均等等。IMC’ 2015[3] 通

过有监督的机器学习算法来解决手动和迭代的调整检测器参数和阈值的难题。多年来，人们提出了数十种异常检测器，用于密切监控设备性能及发现异常。但是部署它们是一个巨大的挑战，需要人工手动和迭代的调整检测器参数和阈值。

该文章通过对多个检测器中的性能数据提取异常特征；然后用特征和标签训练随机森林分类器，自动选择适当的参数和阈值。有以下局限性：1. 有监督学习需要带标签的数据；2. 这种算法受限于训练集中的异常，也就是无法判别未来出现的新的异常；因此，WWW’2018[4] 提出了无监督的机器学习算法，即针对周期性 KPI 数据，使用基于 VAE(变分自动编码器) 的异常检测算法。

2.3.2　基于日志的异常检测

系统产生的日志是非结构化文本，并且有信息量巨大、类型繁多等特点，这就为分析日志带来了许多困难。主要有以下几点挑战：

1. 系统越来越多地由多个组件，尤其是分布式组件构成，使用单个日志文件监视系统变得不可能。交叉异构的日志文件很难分析，特别是当时间戳不同步甚至不存在的时候。2. 最大化日志系统的信息量的同时最小化检测成本。3. 单纯的统计模型并不能提供可行性建议。例如，可以用机器学习模型来发现负载中的异常，CPU 利用率过高，但是无法解释应该怎么处理。CCS’2017[5] 提出了一种基于深度学习的日志异常检测系统——DeepLog。DeepLog 通过以下三步异常检测来综合判断系统异常. 1. 执行路径异常检测。将异常检测问题转换成一个 log key 的多分类问题，使用 LSTM 对日志的 log key 序列建模，自动从正常的日志数据中学习正常的模式并且由此来判断系统异常。同时，LSTM 可以增量式地调整模型参数，以便适应随着时间推移而出现的新日志文件。2. 参数和性能异常检测。有时候系统虽然是按照正常操作步骤执行的，但是记录的日志中的参数是不正常的，比如延迟比正常要大，这种情况也属于异常。3. 工作流异常检测。虽然工作流模型在异常检测的有效性上不如 LSTM 模型，但是工作流模型可以可视化地帮助运维工程师在发现异常后找出异常的原因。其中长短期记忆（Long short-term memory, LSTM）是一种特殊的 RNN，主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题。相比普通的 RNN，LSTM 能够在更长的序列中有更好的表现。

2.4　异常检测算法介绍

本节将介绍在异常检测领域主流的一些算法，根据所依赖的技术原理的不同，将这些算法分为了基于统计、基于分类、基于聚类、基于深度学习的异常检测算法。

2.4.1　基于统计的异常检测算法

早期的异常检测方法往往基于统计与概率模型，也就是假设-检验的方法。首先对数据的分布做出假设，然后找出假设下所定义的“异常”，因此往往使用极值分析或假设检验。比如对最简单的一维数据假设服从正态分布，然后将距离均值某个范围以外的点当做异常点。推广到高维后，假设各个维度相互独立。这类方法的好处速度一般比较快，但是因为存在很强的“假设”，效果不一定很好。在时间序列异常检测领域，最常见的基于统计的算法为 ARIMA，即差分自回归移动平均模型 [8]。

2.4.2　基于分类的异常检测算法

这类方法通常是将异常检测看成是数据不平衡下的分类问题。常用分类算法有朴素贝叶斯、逻辑回归、支持向量机等。以支持向量机 (SVM) 为例，SVM 在集成学习和神经网络之类的算法没有表现出优越性能前，基本占据了分类模型的统治地位。目前由于互联网数据规模的急剧膨胀,SVM 无法很好的处理海量样本，热度有所下降，但是仍然是一个常用的机器学习算法。在异常检测领域中，当异常值远远少于正常值时，可以用 One-Class SVM。当异常值较多即正负样本均衡时，适用于普通的二分类 SVM。可以根据样本情况灵活调整。One-Class SVM 的输入为不包含异常值的“干净”数据，试图求得高维空间的一个超球面，以最小的半径将训练集中的数据包起来。新来的待测数据映射到高维空间后，如果落在这个超球面之外，则认为它是一个异常值。

2.4.3　基于聚类的异常检测算法

聚类算法通常是基于距离/密度发现异常点。基于距离/密度的异常点检测方法的关键步骤在于给每个数据点都分配一个离散度，其主要思想是：针对给定的数据集，对其中的任意一个数据点，如果在其局部邻域内的点都很密集，那么认为此数据点为正常数据点，而异常点则是距离正常数据点最近邻的点都比较远的数据点。通常有阈值进行界定距离的远近。异常检测领域下常用的聚类算法有 k-means、LOF、孤立森林、高斯混合模型 [20] 等。

2.4.4　基于深度学习的异常检测算法

随着深度学习的兴起，越来越多的学者尝试用深度学习算法来进行异常检测，尤其是针对时间序列数据，深度学习模型往往表现出惊人的效果。常用的深度学习算法为变分编码器、神经网络 [6][14]、生成对抗网络、LSTM[17]、

RNN[3][4][10][12][13][15] 等。以变分自动编码器 (Variational Auto-Encoder)[5] 为例，其利用自编码器的重构误差和局部误差，针对时间序列的异常检测的场景，达到了很好的效果。

2.5　异常检测领域开源数据集介绍

数据集主要由 KDDCUP99, CICIDS 等。

2.6　异常检测算法对比

2.7　现有异常检测算法存在的问题

第 3 章　基于深度学习的时间序列异常检测算法研究

3.1　引言

3.2　循环神经网络原理

第 4 章　引用文献的标注

模板支持 BibTeX 和 BibLaTeX 两种方式处理参考文献。下文主要介绍 BibTeX 配合 natbib 宏包的主要使用方法。

4.1　顺序编码制

在顺序编码制下，默认的 `\cite` 命令同 `\citep` 一样，序号置于方括号中，引文页码会放在括号外。统一处引用的连续序号会自动用短横线连接。

`\cite{zhangkun1994}` ⇒ [4]

`\citet{zhangkun1994}` ⇒ 张昆 等[4]

`\citep{zhangkun1994}` ⇒ [4]

`\cite[42]{zhangkun1994}` ⇒ [4]42

`\cite{zhangkun1994,zhukezhen1973}` ⇒ [4-5]

也可以取消上标格式，将数字序号作为文字的一部分。建议全文统一使用相同的格式。

`\cite{zhangkun1994}` ⇒ [4]

`\citet{zhangkun1994}` ⇒ 张昆 等 [4]

`\citep{zhangkun1994}` ⇒ [4]

`\cite[42]{zhangkun1994}` ⇒ [4]^{42}

`\cite{zhangkun1994,zhukezhen1973}` ⇒ [4-5]

4.2　著者-出版年制

著者-出版年制下的 `\cite` 跟 `\citet` 一样。

`\cite{zhangkun1994}` ⇒ 张昆 等 (1994)

`\citet{zhangkun1994}` ⇒ 张昆 等 (1994)

`\citep{zhangkun1994}` ⇒ (张昆 等, 1994)

`\cite[42]{zhangkun1994}` ⇒ (张昆 等, 1994)^{42}

`\citep{zhangkun1994,zhukezhen1973}` ⇒ (张昆 等, 1994; 竺可桢, 1973)

注意，引文参考文献的每条都要在正文中标注[4-37]。

参考文献

[1] 中国互联网信息中心. 中国互联网络发展状况统计报告 [Z].

[2] Ahmed M, Mahmood A N, Hu J. A survey of network anomaly detection techniques[J]. Journal of Network and Computer Applications, 2016, 60: 19-31.

[3] Hawkins D M. Identification of outliers: volume 11[M]. Springer, 1980.

[4] 张昆, 冯立群, 余昌钰, 等. 机器人柔性手腕的球面齿轮设计研究 [J]. 清华大学学报: 自然科学版, 1994, 34(2): 1-7.

[5] 竺可桢. 物理学论 [M]. 北京: 科学出版社, 1973: 56-60.

[6] Dupont B. Bone marrow transplantation in severe combined immunodeficiency with an unrelated mlc compatible donor[C]// White H J, Smith R. Proceedings of the third annual meeting of the International Society for Experimental Hematology. Houston: International Society for Experimental Hematology, 1974: 44-46.

[7] 郑开青. 通讯系统模拟及软件 [D]. 北京: 清华大学无线电系, 1987.

[8] 姜锡洲. 一种温热外敷药制备方案: 中国, 88105607.3[P]. 1980-07-26.

[9] 中华人民共和国国家技术监督局. GB3100-3102. 中华人民共和国国家标准-量与单位 [S]. 北京: 中国标准出版社, 1994.

[10] Merkt F, Mackenzie S R, Softley T P. Rotational autoionization dynamics in high rydberg states of nitrogen[J]. J Chem Phys, 1995, 103: 4509-4518.

[11] Mellinger A, Vidal C R, Jungen C. Laser reduced fluorescence study of the carbon monoxide nd triplet rydberg series - experimental results and multichannel quantum defect analysis[J]. J Chem Phys, 1996, 104: 8913-8921.

[12] Bixon M, Jortner J. The dynamics of predissociating high Rydberg states of NO[J]. J Chem Phys, 1996, 105: 1363-1382.

[13] 马辉, 李俭, 刘耀明, 等. 利用 REMPI 方法测量 BaF 高里德堡系列光谱 [J]. 化学物理学报, 1995, 8: 308-311.

[14] Carlson N W, Taylor A J, Jones K M, et al. Two-step polarization-labeling spectroscopy of excited states of Na2[J]. Phys Rev A, 1981, 24: 822-834.

[15] Taylor A J, Jones K M, Schawlow A L. Scanning pulsed-polarization spectrometer applied to Na2[J]. J Opt Soc Am, 1983, 73: 994-998.

[16] Taylor A J, Jones K M, Schawlow A L. A study of the excited 1Σg+ states in Na2[J]. Opt Commun, 1981, 39: 47-50.

[17] Shimizu K, Shimizu F. Laser induced fluorescence spectra of the a 3Πu–X 1Σg+ band of Na2 by molecular beam[J]. J Chem Phys, 1983, 78: 1126-1131.

[18] Atkinson J B, Becker J, Demtröder W. Experimental observation of the a 3Πu state of Na2[J]. Chem Phys Lett, 1982, 87: 92-97.

[19] Kusch P, Hessel M M. Perturbations in the a 1Σu+ state of Na2[J]. J Chem Phys, 1975, 63: 4087-4088.

[20] 广西壮族自治区林业厅. 广西自然保护区 [M]. 北京: 中国林业出版社, 1993.

[21] 霍斯尼. 谷物科学与工艺学原理 [M]. 李庆龙, 译. 2 版. 北京: 中国食品出版社, 1989: 15-20.

[22] 王夫之. 宋论 [M]. 刻本. 金陵: 曾氏, 1865（清同治四年）.

[23] 赵耀东. 新时代的工业工程师 [M/OL]. 台北: 天下文化出版社, 1998[1998-09-26]. http://www.ie.nthu.edu.tw/info/ie.newie.htm.

[24] 全国信息与文献工作标准化技术委员会出版物格式分委员会. GB/T 12450-2001 图书书名页 [S]. 北京: 中国标准出版社, 2002.

[25] 全国出版专业职业资格考试办公室. 全国出版专业职业资格考试辅导教材: 出版专业理论与实务 • 中级 [M]. 2014 版. 上海: 上海辞书出版社, 2004: 299-307.

[26] World Health Organization. Factors regulating the immune response: Report of WHO Scientific Group[R]. Geneva: WHO, 1970.

[27] Peebles P Z, Jr. Probability, random variables, and random signal principles[M]. 4th ed. New York: McGraw Hill, 2001.

[28] 白书农. 植物开花研究 [M]// 李承森. 植物科学进展. 北京: 高等教育出版社, 1998: 146-163.

[29] Weinstein L, Swertz M N. Pathogenic properties of invading microorganism[M]// Sodeman W A, Jr, Sodeman W A. Pathologic physiology: mechanisms of disease. Philadelphia: Saunders, 1974: 745-772.

[30] 韩吉人. 论职工教育的特点 [C]// 中国职工教育研究会. 职工教育研究论文集. 北京: 人民教育出版社, 1985: 90-99.

[31] 中国地质学会. 地质评论 [J]. 1936, 1(1)–. 北京: 地质出版社, 1936–.

[32] 中国图书馆学会. 图书馆学通讯 [J]. 1957(1)–1990(4). 北京: 北京图书馆, 1957–1990.

[33] American Association for the Advancement of Science. Science[J]. 1883, 1(1)–. Washington, D.C.: American Association for the Advancement of Science, 1883–.

[34] 傅刚, 赵承, 李佳路. 大风沙过后的思考 [N/OL]. 北京青年报, 2000-04-12(14)[2002-03-06]. http://www.bjyouth.com.cn/Bqb/20000412/B/4216%5ED0412B1401.htm.

[35] 萧钰. 出版业信息化迈入快车道 [EB/OL]. (2001-12-19)[2002-04-15]. http://www.creader.com/news/20011219/200112190019.htm.

[36] Online Computer Library Center, Inc. About OCLC: History of cooperation[EB/OL]. 2000 [2000-01-08]. http://www.oclc.org/about/cooperation.en.htm.

[37] Scitor Corporation. Project scheduler[CP/DK]. Sunnyvale, Calif.: Scitor Corporation, 1983.

附录 A　补充内容

附录是与论文内容密切相关、但编入正文又影响整篇论文编排的条理和逻辑性的资料，例如某些重要的数据表格、计算程序、统计表等，是论文主体的补充内容，可根据需要设置。

A.1　图表示例

A.1.1　图

附录中的图片示例（图 A.1）。

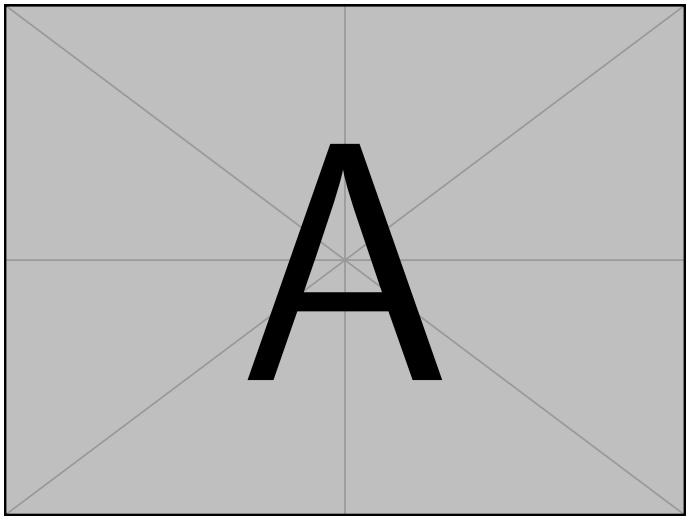

图 A.1　附录中的图片示例

A.1.2　表格

附录中的表格示例（表 A.1）。

A.2　数学公式

附录中的数学公式示例（公式 (A-1)）。

$$\frac{1}{2\pi \mathrm{i}} \int_{\gamma} f = \sum_{k=1}^{m} n(\gamma; a_k) \mathscr{R}(f; a_k) \tag{A-1}$$

表 A.1　附录中的表格示例

文件名	描述
thuthesis.dtx	模板的源文件，包括文档和注释
thuthesis.cls	模板文件
thuthesis-*.bst	BibTeX 参考文献表样式文件
thuthesis-*.bbx	BibLaTeX 参考文献表样式文件
thuthesis-*.cbx	BibLaTeX 引用样式文件

致　谢

衷心感谢导师 ××× 教授和物理系 ×× 副教授对本人的精心指导。他们的言传身教将使我终生受益。

在美国麻省理工学院化学系进行九个月的合作研究期间，承蒙 Robert Field 教授热心指导与帮助，不胜感激。

感谢 ××××× 实验室主任 ××× 教授，以及实验室全体老师和同窗们学的热情帮助和支持！

本课题承蒙国家自然科学基金资助，特此致谢。

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____ 日　期：_____