

基于机器学习的光链路建立中的传输质量预测技术

鄢 然^{1,2}, 郑 豪³, 李 蔚^{3*}

(1. 武汉中原电子集团有限公司, 武汉 430205; 2. 华中科技大学 软件学院, 武汉 430074;

3. 华中科技大学 武汉光电国家研究中心, 武汉 430074)

摘要: 传输质量(QoT)预测在光网络中日趋重要, 机器学习成为今后实现光网络中 QoT 预测的重要手段。提出一种基于机器学习分类器的 QoT 预测技术。通过传输方程生成所需的数据, 用于之后的分类器训练和性能测试, 并仿真验证了 K 最近邻(KNN)、逻辑回归(LR)和支持向量机(SVM)这 3 种常用的分类器的性能。仿真结果表明: 相较于传统的 QoT 估计方法, 基于机器学习的方法在有效地降低计算复杂度的前提下, 还能提供相当高的预测精度, 是一种具有广阔应用前景的 QoT 估计新方案。

关键词: 光网络; 传输质量; 非线性光传输方程; 机器学习; 高斯噪声模型

中图分类号: TN256 文献标识码: A 文章编号: 1002-5561(2020)06-0015-05

DOI: 10.13921/j.cnki.issn1002-5561.2020.06.004

开放科学(资源服务)标识码(OSID):



Transmission quality prediction technology in optical link building based on machine learning

YAN Ran^{1,2}, ZHENG Hao³, LI Wei^{3*}

(1. Wuhan Zhongyuan Electronics, Wuhan 430205, China;

2. School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;

3. Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: The prediction of transmission quality (QoT) is becoming more and more important in optical networks. Machine learning has become an important means to realize the prediction of QoT in optical networks in the future. This paper presents a new technology of QoT prediction based on machine learning classifier. Through the data generated by the transfer equation, it can be used for classifier training and performance testing. The performance of K-Nearest Neighbor (KNN), logistic regression (LR) and support vector machine (SVM) are verified by simulation. The simulation results show that, compared with the traditional QoT estimation method, the machine learning based on method can effectively reduce the computational complexity and provide quite high prediction accuracy. This method is a new QoT estimation scheme with broad application prospects.

Key words: optical network; quality of transmission; nonlinear optical transmission equations; machine learning; Gaussian noise model

0 引言

传统的光网络性能研究侧重于计算网络层参数, 这些参数仅基于网络的可用容量和流量负载^[1]。然而,

在透明光网络的情况下, 光路是否被阻塞不再仅仅由网络层因素决定, 它还受物理层因素(如传输质量(QoT))的影响^[2-5]。因此, 必须在透明光网络中部署新光路之前预测其 QoT。通常的 QoT 涉及的物理层参数包括光信噪比(OSNR)、误码率(BER)和 Q 因子。目前, QoT 预测技术可分为两大类^[6]: 一大类是不使用机器学习的传统技术; 另一大类是使用基于机器学习分类器的技术。其中, 第一类技术还包括 2 种方式: 一种是采用分析模型^[7-9]来估计物理层损伤并提供准确的结果, 计算复杂性高; 另一种是近似公式法^[10], 计算速度快, 但结果预测精度不高, 且近似公式经常引入高

收稿日期: 2019-12-03。

基金项目: 国家重点研发计划项目(2018YFB2200900)资助。

作者简介: 鄢然(1994—), 女, 硕士研究生, 湖北天门人, 主要研究方向为计算机网络、光网络以及机器学习等相关方面。参与国家重点研发计划项目, 针对光网络中光链路传输质量的预测问题进行了分析研究及其仿真验证。

* 通信作者: 李蔚(E-mail: weilie@hust.edu.cn)。



余量导致未充分利用网络资源。基于机器学习分类器的技术与传统技术相比具有许多优点,可满足高精度和实时要求,它们还可以自动预测未建立光路的 QoT^[11-17]。文献[17]中,作者使用了3种分类器:K最近邻(KNN)、随机森林(RF)以及支持向量机(SVM),分别达到81.75%、96.32%和99.15%的分类准确度。文献[6]中,作者采用了RF二进制分类器来预测未建立光路的BER将超过系统阈值的概率。在上述2篇文献中,都是采用高斯噪声(GN)模型来生成训练集和测试集数据。然而,由于GN模型中非线性噪声的估计并不准确,从训练数据获得的分类器性能可能与实际情况存在较大误差。因此,本文提出一种基于机器学习分类器的QoT预测技术,通过求解波分复用(WDM)传输方程来获得人工合成数据以进行进一步的训练和测试。

1 机器学习分类器

分类器结构如图1所示。本文所提出的分类器有以下特征:跨段数、跨段长度、调制格式、比特率和信道发射功率。分类器的预测目标变量是布尔逻辑变量,仅当预测的BER小于阈值BER时才为真,否则为假。本文对KNN、逻辑回归(LR)和SVM这3种常用的机器学习分类器进行分析研究。当预测输出与训练集中给出的标签之间的差异达到停止条件时,分类器的训练阶段已经结束。

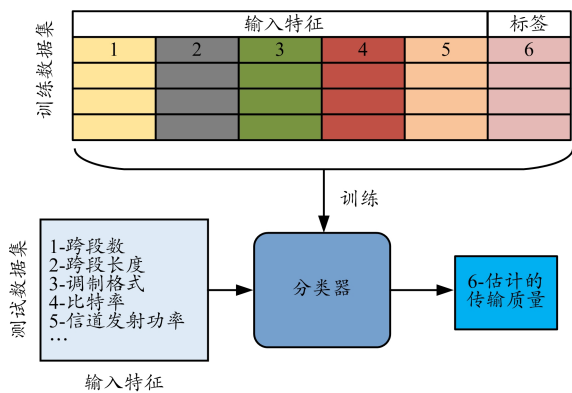


图1 分类器的结构

1.1 KNN

KNN是必不可少的简单分类和回归方法。KNN分类器基于训练数据集对应特征空间的划分。在KNN中,当确定训练数据集、距离度量、K值和分类决策规则时,KNN将会唯一确定一个分类结果。

KNN的3个要素包括距离度量、K值选择和分类决策规则。距离度量包括常见的欧氏距离和其它一些

更加一般化的距离。K值选择是当K值小时,KNN分类器会更复杂;当K值很大时,KNN分类器会更简单。K值的选择反映了近似误差和估计误差之间的折衷,通常通过交叉验证来选择。常用的分类决策规则是多数表决,给定输入测试样本后,找到距离度量最近的K个训练样本中类别的多数,得到测试样本的类别。

1.2 LR

LR来自线性回归,对回归结果执行逻辑函数运算,该运算将范围限制为[0,1]区间,并将损失函数更改为二进制交叉熵损失,从而可以将其用于2个分类问题(通过比较概率值和阈值进行分类)。Logistic回归要求输入标签数据是伯努利分布;而线性回归是任意连续值的回归,它是由统计学家David Cox于1958年首次提出的。

1.3 SVM

SVM是一种基于统计学习理论的新型机器学习算法。它使用结构风险最小化的原理代替经验风险最小化的原理来解决小样本学习的问题。它还使用核函数的思想将非线性空间问题转化为线性空间,从而降低了算法的复杂度。由于其完整的理论基础和良好的学习性能,SVM在解决有限样本、非线性和高维模式识别问题方面具有许多独特优势,现已成为机器学习领域的热点之一,并成功应用于许多领域。

2 仿真设置和结果分析

2.1 仿真系统设置

考虑到获得真实现场监测数据的难度,本文使用人工合成数据来训练和测试所提出的基于机器学习的分类器。为了构建由人工合成数据组成的知识库(KB),有必要计算与不同输入特征的变化值相对应的BER。与使用GN模型^[10]的近似公式方法相比,通过分步傅里叶方法(SSFM)求解WDM传输方程来获得BER可以提供相对准确的结果。此外,虽然求解WDM传输方程的计算复杂度高,但人工合成数据的生成可以看作是分类器训练的初步准备,花费的时间并不是未来分类器预测中的性能指标。因此,本文使用相对耗时的解决方案,通过SSFM求解WDM传输方程来获得BER^[7,8]。

为了验证使用WDM传输方程产生的人工合成数据的性能,本文基于Matlab 2017a使用GN模型生成人工合成数据以进行比较。KB由2个部分组成:第一部分由GN模型生成;第二部分由WDM传输方程生成。使用的仿真系统结构如图2所示。

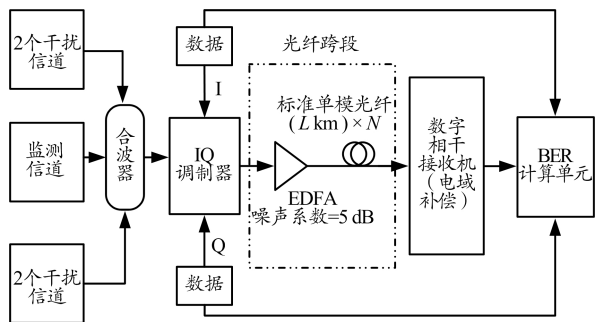


图2 仿真系统结构

该仿真系统由5个发射器、 N 个单模光纤、 N 个掺铒光纤放大器(EDFA)的光学链路、数字相干接收机(集成了电域补偿功能)和BER计算单元组成。EDFA可以精确地补偿光纤损耗,但也会增加噪声。在不失一般性的情况下,假设中心信道是监测信道,研究其性能和交叉相位调制(XPM)效应,而其它4个信道是干扰信道。光纤参数包括群速度色散系数、光纤损耗和非线性参数,分别用 β 、 α 和 γ 表示。接收端执行相干解调和电域补偿以确定探测信道的BER。

对于WDM传输方程仿真产生的KB的第二部分,每个实例由总光路长50~5000 km,跨度长50 km、80 km和100 km,跨段数为1~50,信道发射功率从-10~4 dBm变化,以2 dBm为间隔,以及3种调制格式(QPSK、16QAM和64QAM)等特征组成。本文假设EDFA的噪声系数为5 dB,监测信道的中心频率为193.4 THz,通道间隔设置为50 GHz,噪声带宽为32 GHz,使用的符号率为32 Gbaud。根据ITU-T G.975.1建议的前向纠错(FEC)标准,将Pre-FEC BER与阈值($T=4 \times 10^{-3}$)进行比较,得到的这部分KB包含3600种不同光路和相应标签。其中,有1118个 $BER < T$ 和2482个 $BER > T$ 的实例。

对于GN模型生成KB的第一部分,每个实例由总光路长80~7500 km,跨度长80 km、100 km、120 km和150 km,跨度数为1~50,信道发射功率从-10~4 dBm变化(以2 dBm为间隔),以及3种调制格式(PM-BPSK、PM-QPSK和16QAM)等特征组成。本文将EDFA的噪声系数设置为5 dB,通道间隔设置为50 GHz,噪声带宽为32 GHz,符号率为32 Gbaud。首先,GN模型用于粗略估计非线性噪声,并且通过将非线性噪声与放大器自发辐射噪声(ASE)组合来计算OSNR;然后,通过OSNR和BER之间的映射关系获得相应的BER。基于GN模型的数据集总共有4800个实例。其中, $BER < T$ 和 $BER > T$ 的数量分别为1692和3108。

2.2 性能分析

为测试训练分类器的性能,本文将KB分为3种情况,如图3所示。情况I:训练数据集是从KB的WDM传输方程中随机抽取的2520个实例获得。测试数据集由剩余的非线性薛定谔方程(NLSE)数据1080个实例组成。情况II:训练数据集完全由从KB的GN模型中随机提取的2520个实例生成,并且测试数据集与情况I保持不变。情况III:随机抽取情况I和情况II训练数据集的50%,并将它们合并成新的训练数据集,而测试数据集保持与情况I和情况II相同。在3种情况下测试数据集是由传输方程生成的相同数据集,这是因为由传输方程生成的数据与实际现场数据足够相似,能够准确地测量训练分类器的性能。

然后,使用KNN、LR和SVM 3种机器学习分类器对KB划分的3种情况分别进行分类,比较它们的性能,其工作流程图如图4所示。本文使用训练好的最佳参数分类器进行光路BER的预测。如果光路预测的 $BER < T$,则对应的布尔逻辑变量等于1,否则为0,这是分类器预测的标签。标签等于1意味着它是“好的QoT”;而标签等于0意味着它是“差的QoT”。对于KB中数据的实际标签,我们也采用相同的逻辑,即 $BER < T$ 的数据样本的实际标签赋值为1, $BER > T$ 的数据样本,实际标签赋值为0。

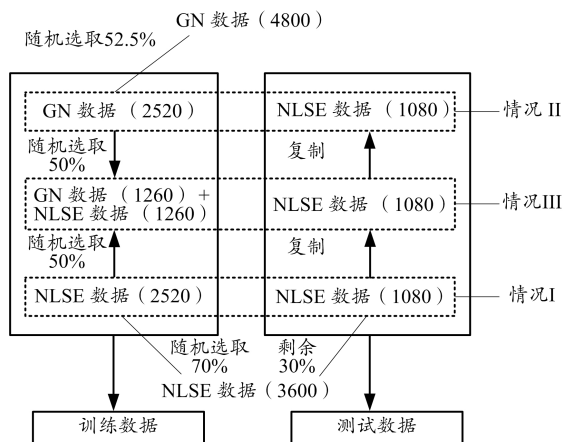


图3 KB划分的3种情况

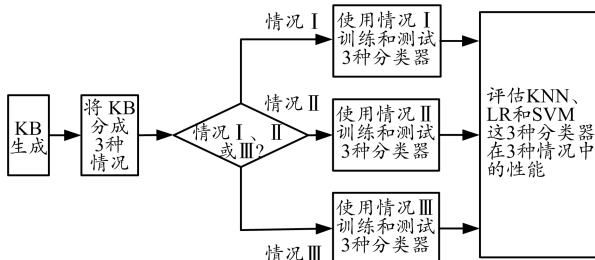


图4 工作流程图

本文采用的分类器对 3 种不同数据分类情况的混淆矩阵如图 5 所示。使用 2 个度量来评估分类器:分类准确率和假正率。假正率(预测标签为“好的 QoT”而实际标签为“差的 QoT”)的实例可用于进一步细化考察分类器的预测性能,如表 1 所示。可以看出,当数据集为情况 I 时,3 种分类器可以达到最高的分类准确率,分别为 97.87%、89.63%和 99.17%;当数据集为情况 II 时,3 种分类器的分类准确率最低,分别为 81.76%、82.5%和 86.85%;当数据集为情况 III 时,3 种分类器的分类准确率在前两者之间,分别为 96.94%、84.91%和 98.06%。通过观察 3 种情况下假正率与分类准确率之间的关系,很容易发现假正率与分类准确率呈负相关。这表明分类器的准确率越高,分类器的假

正率就越低。因此,具有最高分类精度的分类器将具有最佳的整体性能。

由于 GN 模型对非线性噪声的估计并不准确,考虑到 WDM 传输方程相较于 GN 模型能够更准确地对传输过程中受到的非线性损伤进行模拟,即 WDM 传输方程产生的数据相较于 GN 模型产生的数据要更加接近实际光路中受到的真实非线性损伤的数据。由于获取大量的真实现场实测数据存在极大的困难,因此在 3 种情况下,本文使用的测试集都是用 WDM 传输方程产生的数据组成,并使用测试集对训练出的分类器的性能进行评判,使其性能评价结果也更加合理可靠。在这种测试集测试下,情况 II 中采用 GN 模型数据组成的训练集训练出的 3 种分类器的预测准确率都较低;而采用 GN 模型数据和 WDM 传输方程数据各半混合的训练集数据的情况 III,预测准确率得到了提升;情况 I 中,当训练数据集完全由 WDM 传输方程数据组成时,分类器的预测准确率达到最高。预测结果反映了 GN 模型对非线性损伤的估计确实不准确,从而导致了训练好的分类器对测试数据集的分类准确率偏低。因此,由 WDM 传输方程数据训练得到的分类器会比 GN 模型数据训练得到的分类器更加可靠。通过比较情况 I 和情况 II 中获得的结果可知,SVM 在 2 种情况下都是最好的分类器,但 KNN 却有所不同。情况 I 中,KNN 的分类准确率略低于 SVM,但远高于 LR。情况 II 中,KNN 不如 LR 准确。考虑到情况 I 训练的分类器比情况 II 训练的分类器更可靠。因此,3 个分类器的真正优先级应该是 SVM>KNN>LR(情况 I)。

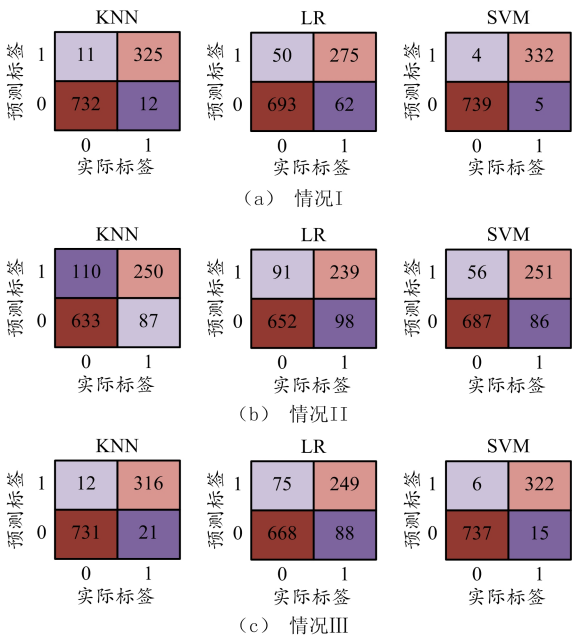


图 5 混淆矩阵

表 1 分类器的预性能

情况	分类器类型	分类器的预测性能	
		准确率/%	假正率/%
I	KNN	97.87	1.02
	LR	89.63	4.63
	SVM	99.17	0.37
II	KNN	81.76	10.19
	LR	82.5	8.43
	SVM	86.85	5.19
III	KNN	96.94	1.11
	LR	84.91	6.94
	SVM	98.06	0.56

3 结束语

本文分析了用于预测未建立光路的 QoT 的 3 种常用的监督学习分类器。通过比较了 KNN、LR 和 SVM 这 3 种分类器在 3 种不同数据集情况下的性能可知,当训练数据集由 WDM 传输方程产生的数据组成时,KNN、LR 和 SVM 的分类准确率是分别为 97.87%、89.63%和 99.17%;而当训练数据集由 GN 模型生成的数据组成时,分类准确率分别为 81.76%、82.5%和 86.85%。

因此,由 WDM 传输方程数据训练得到的分类器会比 GN 模型数据训练得到的分类器更加可靠。3 种分类器的真正优先级应该是 SVM>KNN>LR。未来的工作包括使用优化方法来减小 KB 大小并使用实际网络收集的 BER 数据来训练分类器。

参考文献:

- [1] KULKARNI P, TZANAKAKI A, MACHUKA C M, et al. Benefits of Q-factor based Routing in WDM Metro Networks [C]// Proceedings of European Conference on Optical Communication (ECOC 2005), September 25–29, 2005, Glasgow, UK. London: IET, 2005: 1–2.
- [2] RAMAMURTHY B, DATTA D, FENG H, et al. Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks[J]. Journal of Lightwave Technology, 1999, 17(10): 1713–1723.
- [3] PACHNICKE S, PASCHENDA T, KRUMMRICH P M. Physical Impairment Based Regenerator Placement and Routing in Translucent Optical Networks [C]// Proceedings of Optical Fiber Communication Conference (OFC 2008), February 24–28, 2008, San Diego, USA. Piscataway: IEEE, 2008: 1–3.
- [4] MARKIDIS G, SYGLETOS S, TZANAKAKI A, et al. Impairment-Constraint-Based Routing in Ultralong-Haul Optical Networks With 2R Regeneration [J]. IEEE Photonics Technology Letters, 2007, 19 (6): 420–422.
- [5] POLITI C T, ANAGNOSTOPOULOS V, MATRAKIDIS C, et al. Physical Layer Impairment Aware Routing Algorithms Based on Analytically Calculated Q-Factor[C]// Proceedings of Optical Fiber Communication Conference (OFC 2006), March 5–10, 2006, Anaheim, USA. Piscataway: IEEE, 2006: 1–3.
- [6] BARLETTA L, GIUSTI A, ROTTONDI C, et al. QoT Estimation for Unestablished Lighpaths using Machine Learning [C]// Proceedings of Optical Fiber Communication Conference (OFC 2017), March 19–23, 2017, Los Angeles, USA. Piscataway: IEEE, 2017: 1–3.
- [7] LEIBRICH J, ROSENKRANZ W. Efficient numerical simulation of multichannel WDM transmission systems limited by XPM [J]. IEEE Photonics Technology Letters, 2003, 15(3): 395–397.
- [8] TAO Z, YAN W, LIU L, et al. Simple Fiber Model for Determination of XPM Effects[J]. Journal of Lightwave Technology, 2011, 29(7): 974–986.
- [9] SHAO J, LIANG X, KUMAR S. Comparison of Split-Step Fourier Schemes for Simulating Fiber Optic Communication Systems [J]. IEEE Photonics Journal, 2014, 6(4): 1–16.
- [10] POGGIOLINI P, BOSCO G, CARENA A, et al. The GN-Model of Fiber Non-Linear Propagation and its Applications[J]. Journal of Lightwave Technology, 2014, 32(4): 694–721.
- [11] MO W, HUANG Y, ZHANG S, et al. ANN-Based Transfer Learning for QoT Prediction in Real-Time Mixed Line-Rate Systems [C]//Proceedings of Optical Fiber Communication Conference (OFC 2018), March 11–15, 2018, San Diego, USA. Piscataway: IEEE, 2018: 1–3.
- [12] MATA J, MIGUEL I, DURÁN R J, et al. Supervised Machine Learning Techniques for Quality of Transmission Assessment in Optical Networks [C]//Proceedings of International Conference on Transparent Optical Networks (ICTON 2018), July 1–5, 2018, Bucharest, Romania. Piscataway: IEEE, 2018: 1–4.
- [13] MORAIS R M, PEDRO J. Evaluating Machine Learning Models for QoT Estimation [C]//Proceedings of International Conference on Transparent Optical Networks (ICTON 2018), July 1–5, 2018, Bucharest, Romania. Piscataway: IEEE, 2018: 1–4.
- [14] BOUDA M, ODA S, AKIYAMA Y, et al. Demonstration of Continuous Improvement in Open Optical Network Design by QoT Prediction using Machine Learning [C]//Proceedings of Optical Fiber Communication Conference (OFC 2019), March 3–7, 2019, San Diego, USA. Piscataway: IEEE, 2019: 1–3.
- [15] PANAYIOTOU T, SAVVA G, SHARIATI B, et al. Machine Learning for QoT Estimation of Unseen Optical Network States [C]//Proceedings of Optical Fiber Communication Conference (OFC 2019), March 3–7, 2019, San Diego, USA. Piscataway: IEEE, 2019: 1–3.
- [16] AZZIMONTI D, ROTTONDI C, TORNATORE M. Using Active Learning to Decrease Probes for QoT Estimation in Optical Networks [C]//Proceedings of Optical Fiber Communication Conference (OFC 2019), March 3–7, 2019, San Diego, USA. Piscataway: IEEE, 2019: 1–3.
- [17] ALADIN S, TREMBLAY C. Cognitive Tool for Estimating the QoT of New Lightpaths[C]// Proceedings of Optical Fiber Communication Conference(OFC 2018), March 11–15, 2018, San Diego, USA. Piscataway: IEEE, 2018: 1–3.

新书推荐

《高速可见光通信关键技术》是国家重点研发计划项目“可见光通信关键技术及系统研发”研究团队的集体成果,是国之重器出版工程系列书之一。作者迟楠是该领域的知名专家,多次得到中央电视台、凤凰台和英国 BBC 等媒体的报道。

该书主要介绍了基于 LED 的高速可见光通信的技术原理。首先,给出了可见光通信的基本概念,对其发展历史进行了追溯,同时对其研究趋势进行了展望;其次,分别从高速可见光的系统结构、信道建模、调制技术、均衡技术和编码技术、复用技术和新材料技术等方面具体介绍了实现高速可见光通信所采用的先进技术和关键算法;接着,介绍了高速可见光通信通信系统实验,给出了本研究团队基于本书介绍的技术理论基础之上的实验成果;最后,对高速可见光通信技术的未来进行了展望。

本书适合从事通信领域尤其是可见光通信研究的工程技术人员,以及高等院校通信工程等相关专业的研究生和教师阅读。

