

TD-LTE 网络中大气波导干扰的分析与预测

Analysis and Prediction of Atmospheric Duct Interference in TD-LTE Networks

孙天宇/SUN Tianyu^{1,3}
周婷/ZHOU Ting²
杨旸/YANG Yang¹

(1. 中国科学院上海微系统与信息技术研究所, 上海 200050;
2. 中国科学院上海高等研究院, 上海 201210;
3. 中国科学院大学, 北京 100049)
(1. Shanghai Institute of Microsyst & Information Technology, Chinese Academy of Sciences, Shanghai 200050, China;
2. Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2018) 02-0019-006

摘要: 使用来自于江苏移动的实时网络侧数据来分析大气波导干扰(ADI)的特征, 同时结合网络侧数据与气象数据, 使用两种机器学习方法对 ADI 强度进行预测, 并相互比较。仿真结果表明: 使用机器学习可以获得不错的 ADI 预测效果, 当训练样本达到 40 000 条时, 准确率与召回率分别可以达到 72% 与 75% 以上。

关键词: 时分复用长期演进(TD-LTE); 大气波导; 机器学习; 干扰预测

Abstract: In this paper, the big data of network-side from the current operated network of China Mobile is used to analyze the characteristics of atmospheric duct interference (ADI). Combining network side data with meteorological data, two machine learning methods are used to predict the ADI intensity, and are compared with each other. The simulation results show that machine learning can achieve good ADI prediction effect. When the training sample reaches 40 000, the accuracy and recall rate can reach 72% and 75% respectively.

Keywords: time division-long term evolution (TD-LTE); atmospheric ducts; machine learning; interference prediction

大气波导是一种发生在低层大气中的物理现象, 在大气波导中电磁波反射系数随高度呈现负梯度, 使得电磁信号被约束在大气波导层内传播, 并比在非波导层中经历更少的衰减。1963年, BEAN 和 DUTTON^[1] 首先设定了大气电磁反射率的公式并证明了大气波导的存在。

文献[2]中作者指出波导层可能导致一系列的影响, 比如: 传输的损失、探空雷达的高度误差等。在雷达系统中, 波导层的存在会减少探测范围^[3]。表面波导会导致气象雷达自动降雨率计算的错误^[4]。在文献[5]中, ORAIZI 和 HOSSEINZADEH 研究了大气波导对于正交频分复用(OFDM)数字广播系统的影响, 他们论证了大气波导会增加电磁波的强度从而产生

干扰并且降低频率复用距离。

一些迹象表明: 大气波导会在现行的大规模时分双工(TDD)网络中造成干扰, 譬如时分同步码分多址(TD-SCDMA)、时分复用长期演进(TD-LTE)甚至 5G 的 TDD 网络。这是因为对于 TDD 系统而言, 上行和下行传输使用的是相同的频率。来自其他基站的下行信号可能对本地的上行信号产生同频干扰(CCI), 也就是所谓的上下行干扰。所以, TDD 网络必须在上下行信号转换之间设计出一段空白时隙作为保护时隙(GP)来避免上下行干扰。但是在大气波导发生的情况下, 来自远端基站的下行信号传播时较正常情况下会经历很少的衰减, 从而干扰本地的上行信号。我们用 ADI 来表示这种大气波

导干扰。目前在业界 CCI 消除技术已经被广泛的研究, 而最广泛使用的技术就是小区内干扰消除协调(ICIC)。文献[6]中作者提出了一种基于时域 ICIC 的功率控制方法。文献[7]中作者提出一种机制, 可以有效消除全球微波互联接入(WiMAX)系统中小区边界用户的 CCI。但上述这些方法大都侧重于用户侧的 CCI 消除, 并不适用于 ADI 的场景。TDD 系统中的大气波导干扰吸引了来自学术和工业界的大量关注。通过来自中国移动的实时网络侧数据, 我们发现在江苏省徐州市, ADI 干扰最严重的情况下, 高达 27.6% 的小区收到的上行干扰大于 -90 dBm, 此时干扰完全淹没了有用信号导致小区内无法正常通信。ADI 严重影响了 TD-

收稿日期: 2018-01-15
网络出版日期: 2018-03-28

LTE 系统的性能,因此 ADI 问题亟需被分析与解决。

目前业界存在 2 种主流的方法来检测与估计大气波导:(1)使用探空雷达测量温度、气压等物理量并根据公式来计算修正后的大气电磁折射率^[2];(2)使用海面杂波雷达预测海洋电磁折射扇区面^[6]。方法 1 测量难度大而且花费较高,而方法 2 只适用于海洋场景。此外,这些方法只关注了大气波导本身,一些对于 ADI 关键问题的解答,譬如 ADI 的发生时间或如何鉴别及预测 ADI,依旧没有被解答或被深入研究。除了对干扰的分析,对于运营商来说,预测 ADI 的发生以提高网络性能是十分重要的。

在这篇文章中,我们创造性地使用来自运营商的实时网络侧数据来进行 ADI 预测。由于 ADI 产生的随机性,我们无法从帧结构上解决。因此,成功地预测 ADI 的发生是解决大气波导干扰的关键。ADI 干扰产生的原因复杂繁多,传统的建模方法并不适用。所以,我们结合了网络侧数据与气象数据来进行预测,并使用了支持向量机(SVM),通过将干扰问题转化为一个分类问题实现预测。此外,考虑到关联特征与预测结果之间的相关性,我们还使用因子分解机(FM)算法,将干扰问题转化为一个线性加二阶组合特征的回归问题。该工作填补了大规模 TDD 网络中 ADI 研究的空白,且帮助移动运营商在大气波导场景下提高网络性能。

1 ADI 的概述

1.1 大气波导现象

文献[2]中表明:大气中异常的电磁信号传播条件来源于大气折射率的变化,折射率如公式(1):

$$N = \frac{77.6}{T} \left(p + \frac{4810e}{T} \right) \quad (1)$$

这里 N 是无量纲的折射系数, T 是绝对温度单位为 K , p 是气压单位为 hPa , e 是水气压单位为 hPa 。考

虑到地球曲率的影响,修正后的折射系数 M 为:

$$M = N + \frac{z}{R_e \times 10^6} \quad (2)$$

z 是距地表的高度, R_e 是地球的半径。在正常情况下 M 随着高度的增加而增加。当 M 具有负梯度的时候,对应的高度上就会产生大气波导,电磁信号就会被捕获而进入这段高度之中。

图 1 展示了几种异常的大气状况。

1.2 ADI 对 TD-LTE 网络的影响

在 ADI 的场景下,信号的传播会经历更少的衰减。从而,来自于超过 GP 最大保护距离外基站的信号有可能对本地的上行信号造成同频干扰。目前,业界对于 TDD 系统同频干扰的研究大都集中在用户侧的下行同频干扰消除。这些技术是建立在准确的信道估计或邻近基站的协同的基础上,并不适用于 ADI 的场景。同时,在无线传感网中广泛使用的定位技术^[9]也不适用于 ADI 场景,这是因为 TDD 系统无法将远端基站与本地用户的信号分离开来,从而我们就无法根据接收信号的强度来确定干扰源。

图 2 展示了大气波导存在时上下行信号的碰撞情况,来自远端基站的下行信号传输超过了 GP 对应的最大距离从而有部分与本地上行信号

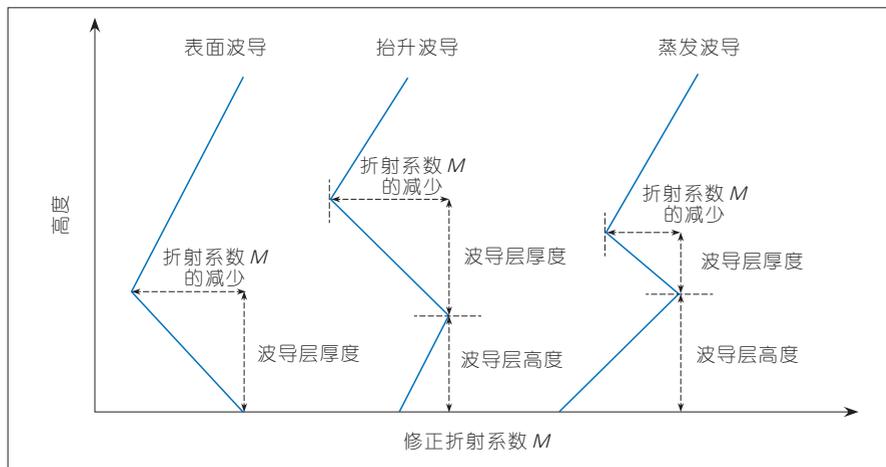
形成重叠,产生同频干扰。ADI 对大规模组网的 TD-LTE 网络会造成严重的影响。根据中国移动的报告,在 2016 年 8 月 25 日,在江苏省徐州市,高达 27.6% 的小区就曾受到了功率大于 -90 dBm 的 ADI 干扰,在这种情况下有用信号完全被干扰淹没,小区的正常通信受到阻碍。因此,有效地预测 ADI 的发生具有很高的研究价值。

1.3 TD-LTE 网络中大气波导干扰的特征

我们使用来自江苏移动 2016 年 5 月 16 日的网络侧数据绘制了干扰地图。图 3 的 a)、b)、c)和 d)分别展示 1:00 am、7:00 am、13:00 pm 和 19:00 pm 时刻全江苏省的 ADI 干扰情况。

在图 3a)中,同频干扰正在逐渐增强,并在图 3b)中达到最强,之后在图 3c)、d)中干扰水平持续降低,并保持稳定,低于 -90 dBm。这些迹象都表明在 1:00 am ~ 7:00 am 这段时间内,ADI 干扰现象很有可能发生。从气象学上可以解释为:在午夜至清晨的这段时间内,地面的温度下降较快,地表的低层大气容易产生一个逆温层,即随着高度的上升大气温度也是上升的。根据公式(1),修正折射率 M 会在该范围内具有一个负梯度,从而导致大气波导层的产生。

此外在图 3b)中的中北部受到的干扰强度会大于南部。考虑到江苏



▲ 图 1 异常大气条件下修正折射系数 M 随高度变化的示例

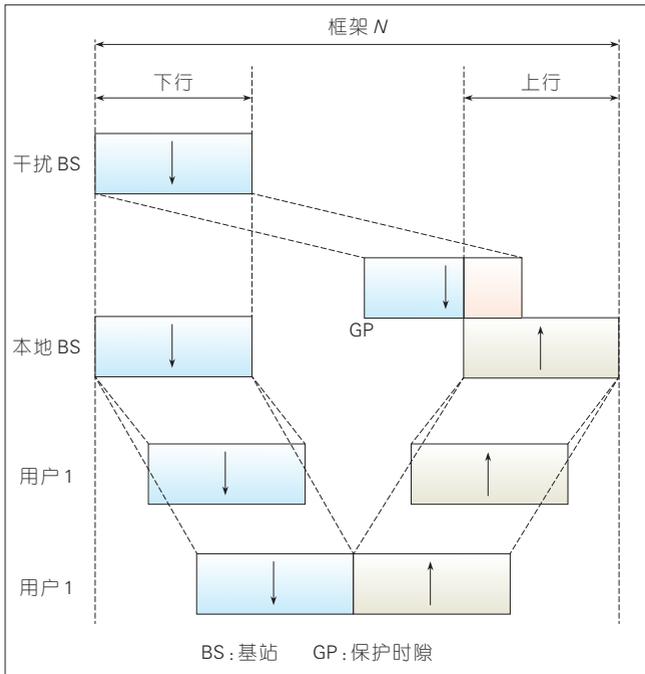


图 2 大气波导存在时上下行信号碰撞情况

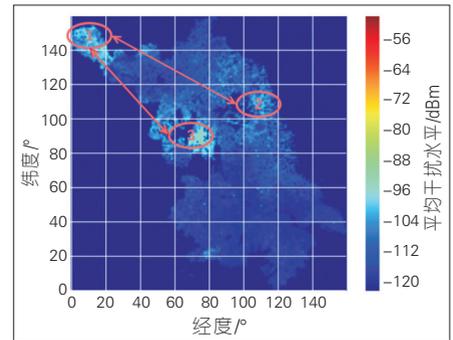


图 4 ADI 互易性质说明

发现 1、2 和 3 号区域的干扰都很强。与 1:00 am 时的情况相比,这 3 个地区的干扰明显都同时增强了。此外图 5 给出了这 3 个地区的受干扰的基站

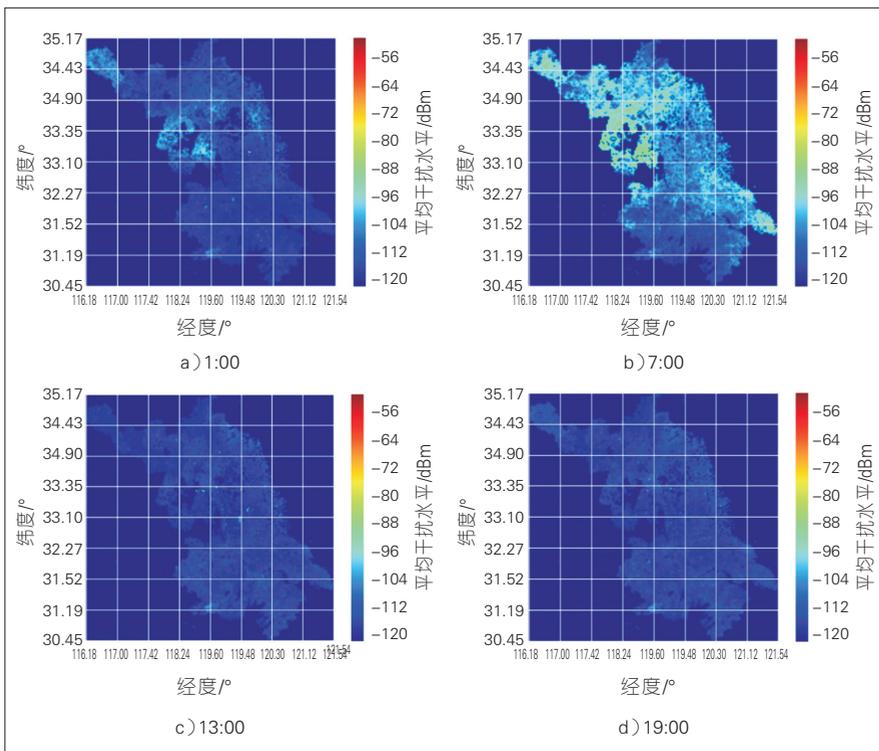


图 3 大气波导干扰地图

省的实际地理情况,中北部农村较多,地势开阔平坦,利于波导层的展开;而在城市之中,高楼较多,干扰信号易被阻断,从而干扰的水平会降低。由此我们可以得出:ADI 具有随

时间变化的特点,而且易于发生在凌晨到清晨这段时间;乡村地区通常有更大的概率受到 ADI 影响。

图 4 展示的是 2016 年 5 月 16 日 2:00 am 的干扰情况。我们可以清楚

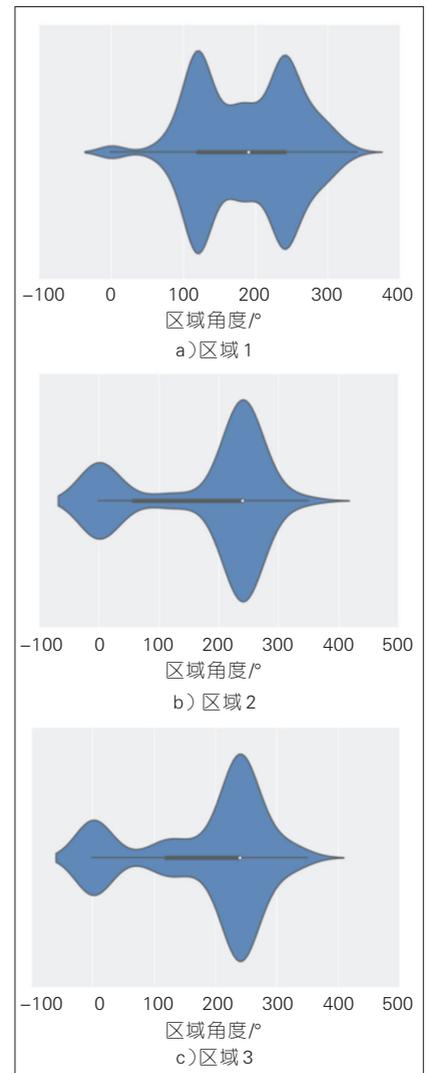


图 5 受 ADI 干扰基站数量在扇区角上的分布

数量在基站扇区角上的分布。我们可以明显发现:区域1的干扰来自于120°和240°的方向,120°正好分别指向区域2和区域3;而区域2和区域3的干扰主要来自240°方向,恰好是区域1的方向。至于区域1来自于240°方向的干扰,极有可能来源于江苏省的西北方向的省份,譬如安徽与河南。图4和图5说明了ADI的方向特性以及互易性质。

上述的种种现象表明:ADI干扰的发生与气象数据和网络侧数据是存在着关联性的,因此我们也可以使用这些数据来对ADI的发生进行有效预测。

2 基于机器学习的ADI预测方法

2.1 基于SVM的分类器训练

SVM是一种二分类的机器学习算法^[10]。分类器意味着当待分类数据输入时,我们可以自动判别它属于哪一类。在我们的问题中,输入是处理过的基站侧数据与气象数据;而输出是1或者-1,分别代表基站是否会被ADI影响。SVM的主要思想是在样本空间中寻找一个鲁棒性最好的划分超平面,使得正负样本间的距离最大。划分超平面可以用 $\omega^T x + b = 0$ 表示。假设我们有训练样本集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 。其中, $y_i \in \{1, -1\}$ 是我们训练样本的标签。在实际中我们根据运营商的需求设定阈值,将干扰大于阈值的样本标记为1,代表此基站受到ADI的影响;将干扰小于阈值的样本标记为-1,代表正常的基站。 x_i 是样本的特征,它由基站侧数据与气象数据通过处理生成。我们使用训练数据集训练出 ω 和 b ,从而获得一个线性分类器: $f(x) = \text{sign}(\omega^T x + b)$ 。

SVM的优化问题可以写为:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \max(1 - y_i(\omega^T x_i + b), 0) \quad (3)$$

其中, C 是常数, $C \sum_{i=1}^m \max(1 - y_i(\omega^T x_i + b), 0)$ 是正则项,我们使用它来避免过拟合与线性不可分的情况。公式(3)是一个二次规划的优化问题,求解它的时间复杂度是 $O(n^2)$,当训练样本量过大会消耗大量的训练时间。文献[11]中,作者提出了一种最小序列优化的方法,避免了对海森矩阵求逆;但该方法侧重于减少空间复杂度,对时间复杂度的优化效果不大。利用文献[12]中提出的乘子交替法(ADMM)框架可以分布式训练,我们可以将时间复杂度降低为 $O(n^2/k^2)$,其中 k 为训练数据的分区数。

2.2 基于因子分解机的回归预测

SVM是一种广义线性模型。显然,在线性模型下,样本中的特征分别单独对最后的标签形成影响,相互之间并不形成关联。而由公式(1)、(2)我们可以发现:ADI的形成受多种气象特征相互影响,即关联特征与标签之间是存在正相关性的。因此,引入两个特征的组合是非常有意义的。

目前主要有2种手段得到组合特征:(1)通过数据分析加人工构造来进行人工特征工程;(2)通过模型做组合特征的学习。方法1需要专家知识,在我们的问题中并不适用;而方法2常用的模型有因子分解机(FM)、神经网络等。

多项式模型是包含特征组合的最直接的一种模型,其表达式如公式(4):

$$y = \omega_0 + \sum_{i=1}^n \omega_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \omega_{ij} x_i x_j \quad (4)$$

此时,我们的问题变为了一个回归问题。对应于2.1节的分类问题,我们令 y_i 表示接收到的ADI干扰强度,单位为dBm。直观地看公式(4),前两项是线性回归模型的表达式,而最后一项则是二阶特征交叉项,即组

合特征项。这类问题可以使用一般的最小二乘规划来求解参数。

然而在ADI问题中,存在着许多类别变量,如基站扇区角、天线下角等。这些一维的特征经过One-Hot编码后维数会大大增加,使得训练数据变得稀疏。在这种情况下,只有当 $x_i x_j$ 同时不为0才能有效,从而学习到的 ω_{ij} 不是充分统计量结果,会导致 ω_{ij} 不准确。

文献[13]中提出的FM方法通过引入隐向量内积来代替交叉项系数,使得 $x_m x_i$ 与 $x_i x_j$ 的参数不再相互独立,从而使得我们可以在样本稀疏的情况下合理地估计参数。改进后的表达式可以写为:

$$y = \omega_0 + \sum_{i=1}^n \omega_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \quad (5)$$

其中, $\langle v_i, v_j \rangle$ 表示两个长度为 k 的向量的内积,即:

$$\langle v_i, v_j \rangle = \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (6)$$

这相当于对我们使用 k 个描述特征的因子 $v_i = (v_{i,1}, \dots, v_{i,k})$,来表示每一维特征分量 x_i 。这样做的好处是:所有满足存在某个 $j \neq i$,使得 $x_i x_j \neq 0$ 的样本都可以用来学习隐向量 v_i ,在很大程度上避免了数据稀疏造成参数估计不准确的影响。

3 仿真结果

3.1 仿真环境

我们通过仿真来验证和比较两种机器学习预测算法的性能,仿真环境配置见表1。

3.2 预测结果

仿真在不同大小的训练集下进行,样本数分别设为2 000、4 000、10 000、20 000、40 000。每次训练完,我们按照训练集20%的大小选取测试集进行测试。我们使用了K-近邻^[14](KNN)算法作为比较基准。在

▼表 1 仿真环境配置

操作系统	ubuntu 16.04.2 LTS
编程语言	python2.7
工具库	scikit-learn v0.20.dev0
气象数据来源	中国气象数据网 http://data.cma.cn/
网络侧数据	来自江苏省移动共计 24 4375 个基站,数据维度包含基站的经纬度、时间、扇区角、频段、天线高度、ADI 干扰的强度
气象数据	包含了气温、气压、水汽压这 3 个式(1)中的关键参数,以及其他气象上的物理量

ADI: 大气波导干扰

KNN 算法中,测试样本的标签由样本空间中离它最近的 k 个样本投票决定。显然 KNN 是一种懒惰的方法,它没有任何的学习过程,只是存储数据并查找。当我们不使用任何学习策略的时候,这显然是一种最为本能的预估方法。

我们使用准确率和召回率来评估最后的训练结果:

$$\text{accuracy} = \frac{n_{\text{correct}}}{n_{\text{all}}}, \text{recall} = \frac{n1_{\text{correct}}}{n1_{\text{all}}} \quad (7)$$

其中, n_{all} 表示总测试样本数, $n1_{\text{all}}$ 表示标记为 1,也就是被干扰的测试样本数。 n_{correct} 和 $n1_{\text{correct}}$ 分别表示总测试样本和标记为 1 的样本中被判断正确的样本数。召回率反映了有多少被干扰的基站会被成功预测出来,对于运营商提高用户体验来说,是一

个重要的指标。

图 6(a)、b) 分别展示了 3 种方法在准确率与召回率上的仿真结果。当样本数较少时, KNN 的性能要大大好于 SVM 与 FM,这是因为训练数据集不充分导致 SVM 与 FM 学习不充分。而随着训练样本数的提升, SVM 与 FM 性能提升迅速,因为它们充分学习了训练数据中的特征;而 KNN 因为本身不具备学习能力,性能增长放缓。我们可以发现:当训练样本数大于 18 000, FM、SVM 的性能都要大大优于 KNN。其次, FM 的性能一直优于 SVM,这正是由于 FM 使用了关联特征,提高了自身的泛化性能。

4 结束语

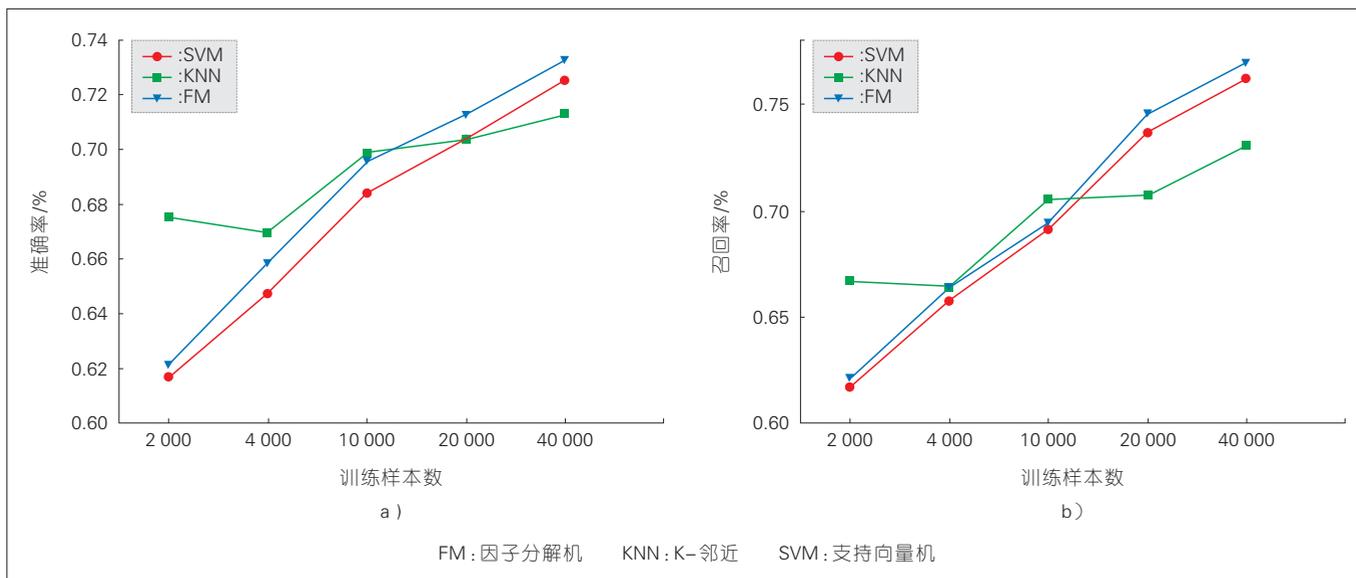
在本文中,我们分析与预测了大规模组网 TD-LTE 系统中的 ADI 问

题。不同于传统气象学中大气波导的检测方法,我们使用了机器学习并结合运营商提供的实时网络侧数据。同时,我们也给出了 ADI 在物理帧结构上的成因,并利用干扰地图等分析手段得出了 ADI 在发生时段与空间分布上的一些特征。

我们首先提出了一种基于 SVM 二元分类器的 ADI 预测方法,创新性地结合了实时网络侧大数据与气象数据。仿真结果表明:随着训练样本数的增长,准确率与召回率都会增长。当训练样本数到达 40 000 时, SVM 的准确率与召回率分别达到 72% 与 75.5%,均显著好于 KNN。

在此基础上,我们提出了一种基于 FM 回归预测 ADI 强度的方法。在线性回归的基础上引入了关联特征,从而提高泛化性能。仿真结果表明:当训练样本数到达 40 000 时, FM 的准确率与召回率分别达到 73.6% 与 76.4%,要优于 SVM。

我们将这篇文章作为 ADI 研究的一个出发点。在未来的工作中仍然有很多的问题亟待解决:(1) ADI 的统计特征与如何对它建模值得研究;(2) 如何精确定位 ADI 的干扰源? 目前来自于运营商的数据在扇区角方面还比较粗糙,只包含 3 个方向的信



▲图 6 各预测方法准确率与召回率性能比较

息,通过场地测量可以获得更为精确的扇区角信息从而帮助进行 ADI 干扰源的定位。

致谢

本文的研究工作和完稿得到了江苏移动研究中心的龚淑蕾和上海无线通信研究中心的张武雄与许晖的帮助和支持,我们表示衷心感谢。

参考文献

- [1] BEAN B R, DUTTON E, FRANK V, et al. A Radio-Meteorological Study[C]// Antennas and Propagation Society International Symposium. USA:IEEE, 1963:56-61
- [2] MENTES S S, KAYMAZ Z. Investigation of Surface Duct Conditions over Istanbul, Turkey [J]. Journal of Applied Meteorology and Climatology, 2007, 46(3): 318-337.DOI: 10.1175/JAM2452.1
- [3] ANDERSON K D. Radar Detection of Low-Altitude Targets in a Maritime Environment [J]. IEEE Transactions on Antennas and Propagation, 1995, 43(6): 609-613.DOI: 10.1109/8.387177
- [4] GERSTOFT P, ROGERS L T, HODGKISS W S, et al. Refractivity from Clutter Using Global Environmental Parameters[C]//IEEE 2001 International Geoscience and Remote Sensing Symposium. USA:IEEE, 2001,6: 2746-2748
- [5] ORAIZI H, HOSSEINZADEH S. The Effect of Atmospheric Duct on Modern OFDM-Based Digital Broadcasting Systems[C]// 33rd European Microwave Conference. USA: IEEE, 2003: 747-750. DOI: 10.1109/EUMA.2003.341061
- [6] LU W, FAN Q, LI Z, et al. Power Control Based Time-Domain Inter-Cell Interference Coordination Scheme in DSCNs[C]//IEEE International Conference on Communications (ICC). IEEE: USA, 2016.DOI: 10.1109/ICC.2016.7511467
- [7] XU W, SEZGINER S. Co-channel Interference Cancellation in Reuse-1 Deployments of WiMAX System[C]//IEEE Wireless Communications and Networking Conference (WCNC). USA: IEEE, 2012:342-346
- [8] LI X H, HE Z S, HE J X, et al. Design of an Ocean Atmospheric Duct Signal Processor [C]// International Symposium on Intelligent Signal Processing and Communication Systems. USA:IEEE, 2010:1-4
- [9] MAO G, ANDERSON B D O, FIDAN B. Path Loss Exponent Estimation for Wireless Sensor Network Localization[J]. Computer Networks, 2007, 51(10): 2467-2483.DOI: 10.1016/j.comnet.2006.11.007
- [10] BURGESS S, VANDENBERGHE L.A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(1):121-167
- [11] PLATT J. A Fast Algorithm for Training Support Vector Machines[J]. Journal of Information Technology, 1998, 2(5):1-28
- [12] BOYD S, PARIKH N, CHU E, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J]. Foundations and Trends in Machine Learning, 2011, 3(1): 1-122
- [13] RENDLE S. Factorization Machines with libFM[J]. Acm Transactions on Intelligent Systems & Technology, 2012, 3(3):1-22. DOI: 10.1145/2168752.2168771
- [14] LAROSE D T. Discovering Knowledge in Data: An Introduction to Data Mining[M]. USA: Wiley-Interscience, 2004

作者简介



孙天宇,中国科学院上海微系统与信息技术研究所硕士在读;主要研究方向是机器学习在无线通信中的各种应用。



王婷,中国科学院上海微系统与信息技术研究所博士毕业,中国科学院上海高等研究院新媒体无线技术研究中心研究员;主要研究方向为异构网络中的资源分配与智能网络;曾获 2015 年中国通信协会技术创新一等奖、上海科学技术进步二等奖;已发表论文 20 余篇,同时有 40 余项专利被批准或正在受理。



杨励,中国科学院上海微系统与信息技术研究所研究员、学位委员会副主任,中国科学院无线传感网与通信重点实验室主任,上海无线通信研究中心主任,上海雾计算实验室联合主任,国际雾计算产学研联盟大中华区主任,中国通信学会学术工作委员会委员等;研究领域包括无线传感器网络(物联网)、新一代移动通信系统(5G)、雾计算与网络技术、开放无线测试验证平台等;承担了国家科技重大专项(03 专项)、国家“863”计划、国家自然科学基金重点等一系列前沿课题研究;获选 IEEE Fellow、中组部“万人计划”、上海市“千人计划”、中科院“百人计划”、上海市“优秀学术带头人”和“领军人才”等荣誉;发表论文 150 余篇,申报专利 80 余项。