



摘要：

- IBM 开发的 IGNITE Defect Classify 应用，采用复杂的认知方法来缩短缺陷周转时间。
- 本文概述了由IGNITE Defect Classify 应用所提供的全新缺陷分类方法，以及如何运用这种方法根据过去的分类历史记录立即得出缺陷分类结果。本文描述了如何运用机器学习、文本预处理和建模等方法来转变和改善传统工作方式，获得从缺陷检测转变为缺陷预防的全新体验。

IBM IGNITE Defect Classify 利用机器智能加速测试

缺陷分类实战经验

Abhishek Mitra
Ilan S Jayaraman
Andrew Williams
IBM 全球企业咨询服务部

简介

“少测，测对”已经成为 IBM 质量和测试服务领域多年来一直坚持的准则。当应用以客户为中心的解决方案时，这种准则可以显著改变客户观念，使其从采用传统的缺陷检测方法转变为缺陷预防和预测方法。本文概述了以该准则为宗旨的一种解决方案。

什么是 IGNITE Defect Classify?

传统上，人们总是通过被动响应的方式来处理缺陷。任何错误分类都可能导致缺陷周转时间的延长，最终影响被测试应用的上市时间。IGNITE Defect Classify 是一种 IBM 认知解决方案，它基于现有缺陷模式，对缺陷起源进行实时分类，从而改变了这种传统观念。这将提高用户找到缺陷根本原因的能力，从而将缺陷快速提交给正确的解决团队。

这能解决什么问题？

当出现缺陷时，有些时候会根据有关缺陷起源的错误假设，将其分配给错误的开发团队或解决人员。缺陷是与代码相关？还是与环境相关？分配人员可以利用自己的洞察或应用经验，将缺陷发送给他们认为可以正确将其解决的小组。

如果分配错误，具有高严重性或重大影响的新缺陷将存放于某个人的收件箱的最底层，等待分析并最终退还给发送者，然后重新分配该缺陷，希望到达正确的响应人员手中。错误分配缺陷会导致解决时间延迟数小时，甚至数天。

如果能够根据自动执行的缺陷基本模式分析对缺陷进行正确分配，便能支持提出者快速采取预防性行动。这将有助于确保缺陷可以通过最



解决方案简介

合适的团队得到解决，从而避免任何不必要的周转环节。

IGNITE Defect Classify 缩短缺陷周转时间

想象一下，来自一家全球保险行业客户的一位测试人员提出了缺陷，并根据个人对这些缺陷的理解，将其分配给了编码人员、环境专员或业务分析师。经过两天的分析之后，编码人员认为这个缺陷是环境问题，应当重新分配给环境专员。然后经过正确的重新分配之后，这个缺陷在几个小时内得到了修复。这种情况太常见了。

如果测试人员从一开始就能正确找到缺陷起源，并将缺陷分配给正确的团队，就不会出现任何时间延迟。

IGNITE Defect Classify 解决了这种问题，并支持完成以下工作：

- 通过认知方式实时找到缺陷起源
- 通过提供置信等级证实分类建议
- 根据提供的反馈，改写分类结果，以供机器学习

IGNITE Defect Classify 引擎能够思考和学习

IGNITE Defect Classify 搭载了智能分类技术^[2]，这是机器学习的一个分支，可对传入的缺陷凭单进行分类并划分优先级，需要极少或无需人工干预。这种方法利用机器学习，根据服务人员的模式和过去的操作，了解过去凭单的分类和分配情况。根据这些信息，它可以创建数据驱动模型，自动对新传入的凭单进行分类。

以下是处理工作流程示例，用于说明如何使用预先训练的 Random Forest 模型处理缺陷凭单，将其分类为代码类型缺陷或非代码类型缺陷。^[1]

数据集描述：

训练数据集是基于一系列已关闭缺陷凭单的文本回答而创建的，这些文本回答提供缺陷概述和详细的缺陷描述以及解决方案详细信息。相应的类别标签（例如，代码类型或非代码类型）由

专家根据解决方案详细信息提供。这一步需要大量经过分类的缺陷。

文本预处理：

文本预处理包括删除缺失数据（例如，没有缺陷概述和输出标签的凭单）。此外，也会删除英文停用词（例如，“but”、“again”、“there”、“about”以及“once”）和特定领域的停用词（例如，测试执行者姓名/电子邮件标识/电话号码等）。测试预处理的输出结果是干干净净的数据集，其中只包含定义缺陷类别的关键词。

快速、准确且智能地预测缺陷分类

如图 1 所示，IGNITE Defect Classify 会使用预先训练的 Random Forest 模型来预测最为可能的缺陷类别。Random Forest 模型是最受欢迎的基于决策树的组合模型。

在此图中，我们假设红色和蓝色圆点分别代表（在特征空间中）存在代码类型缺陷的缺陷凭单和存在非代码类型缺陷的缺陷凭单。Random Forest 模型在训练阶段，学习到了如何划定代码类型缺陷和非代码类型缺陷之间合适的判定边界。当提供新缺陷的详细信息时，该分类器根据已学到的判定边界，找到缺陷的位置，从而预测缺陷的类别。

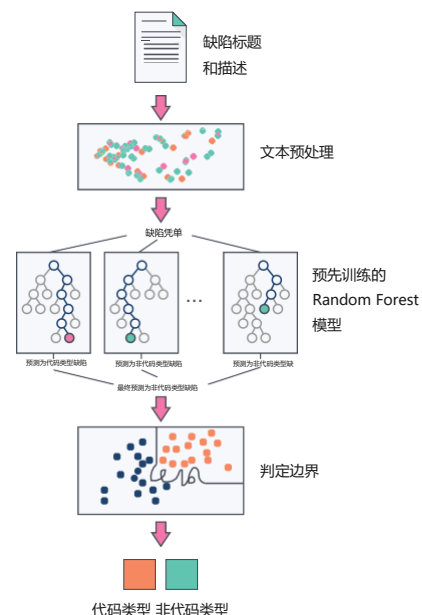


图 1. 使用 Random Forest 分类器进行缺陷分类，实现代码类型和非代码类型的二元缺陷分类

解决方案简介

训练 IGNITE Defect Classify 中的缺陷分类器

作为 IGNITE Defect Classify 的一部分，缺陷分类器需要经过以下几个阶段的训练：

1. 训练阶段：使用由一组文本文档（例如，软件缺陷）以及相应标签构成的数据集训练 Random Forest 算法
2. 预测阶段：利用经过训练的模型来预测未知缺陷的类型。
3. 特征提取：执行检索词频率/反转文档频率向量化处理，从用于训练 Random Forest 分类器的每个预处理文本文档中提取数值特征。

为简化给定示例中的问题，IGNITE Defect Classify 使用了二元分类方法（例如，代码类型/非代码类型）。但是在实际情况中，我们的缺陷凭单数据集包含的标签（例如，代码/设计/要求/测试流程）不止两个，这会导致我们的分类引擎需要处理多类分类问题，最终促使我们将该引擎部署在认知系统中。

为了区别已提出和已关闭缺陷，我们仅通过“缺陷标题”和“缺陷描述”训练了一组模型，用于下述情况：

- 预测已提出缺陷的类别标签
- 通过名为“解决方案详细信息”的其他输入信息训练另一组模型，用于预测已关闭缺陷的类别标签

实际上，这种训练是 IGNITE Defect Classify 的内部机制，它可以智能地使用现有缺陷模式，然后利用这些模式以及其他模型，推荐下一个缺陷应当被发送的解决地点。这会降低缺陷发送错误的可能性，缩短缺陷整体周转时间，减少缺陷修复工作。

解释预测结果

作为 IGNITE Defect Classify 的工作内容，应当对缺陷分类器的结果进行解释说明，以便提取出预测的缺陷分类。

以下是将新缺陷输入到我们的“缺陷类别”分类器中的输出结果示例：

```
{ "defectClass" : "test process" , "accuracy" : 65},
{ "defectClass" : "data" , "accuracy" : 16},
{ "defectClass" : "environment" , "accuracy" : 11},
{ "defectClass" : "requirements" , "accuracy" : 8}}
```

如您所见，该分类器的输出结果是 JSON，其中包含将缺陷归为特定“缺陷类别”的可能性得分（例如，在本示例中“测试流程”的可能性最高）。

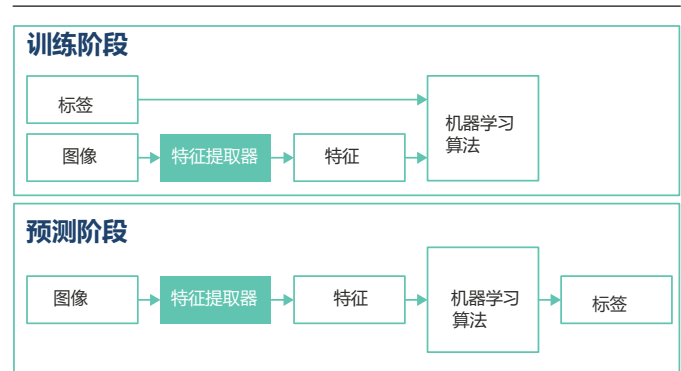


图 2. 分类器训练和测试阶段

总结

通过在测试流程中实施认知缺陷分类技术，可以：

- 显著缩短缺陷周转时间
- 根据缺陷模式分析采取正确的预防措施

本质上，IGNITE Defect Classify 可以快速就缺陷的起源和原因向用户提供通过认知方式得出的建议，确保将缺陷分配给正确的解决人员。这款解决方案可以通过使用自然语言处理器、智能分类以及企业缺陷历史记录，缩短缺陷周转时间，减少缺陷修复工作，这也将改善新业务应用的整体交付速度、提升业务应用性能。

参考资料

[1] Research of Text Categorization Model based on Random Forests, IEEE International Conference on Computational Intelligence and Communication Technology (CICIT), February 2015, <http://ieeexplore.ieee.org/>

[2] Algorithms for Text Categorization: A Comparative Study, World Applied Sciences Journal 22 (9): 1232-1240, 2013, <https://pdfs.semanticscholar.org/b04b/a34d802f4e3f176ad0a8015fccb6ff98addd.pdf>

如有任何疑问，请通过以下方式联系 IBM，我们会为您提供更专业的咨询：

1. 免费咨询电话：400-810-1818 转 2396 (服务时间：9:00-17:00)
2. 填写[需求](#)，提交至 IBM，我们会尽快与您取得联系。



© Copyright IBM Corporation 2018

IBM Corporation
Global Process Services
Route 100
Somers, NY 10589

美国出品
2018 年 2 月

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corp. 在全球许多司法管辖区域的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。Web 站点 ibm.com/legal/copytrade.shtml 上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表。

本文档为自最初公布日期起的最新版本，IBM 可随时对其进行修改。IBM 并不一定在开展业务的所有国家或地区提供所有这些产品或服务。

本文档内的信息“按现状”提供，不附有任何种类的（无论是明示的还是默示的）保证，包括不附有关于适销性、适用于某种特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据其提供时所依据的协议和条款获得保证。



请回收利用
