
垃圾邮件检测的多语言规则

Minh Tuan Vu¹、Quang Anh Tran¹、Frank Jiang²和 Van Quan Tran¹

¹河内大学信息技术学院, 河内, 越南

²新南威尔士大学工程与信息技术学院,
澳大利亚堪培拉

2013 年 4 月 15 日收到; 2014 年 5 月 23 日接受
2014 年 8 月 4 日出版

摘要

在本文中, 我们介绍了一种基于统计规则的方法来为 SpamAssassin 创建规则以检测不同语言的垃圾邮件。还说明了生成和维护多语言规则的理论框架。实验是针对三种语言的数据集进行的, 包括中文、越南语和英语。多语种规则在阈值为2时的真实检测率为89.5%, 报警失败率仅为3.8%, 而单语言规则的真实检测率不超过61%, 报警失败率高达4.9%。

关键词: 垃圾邮件检测, 多语言规则, SpamAssassin, 垃圾邮件, 火腿。

1 介绍

近年来, 打击垃圾邮件的斗争异常激烈。尽管反垃圾邮件技术不断发展, 但垃圾邮件发送者仍在努力寻找新策略, 以帮助向世界各地的电子邮件用户发送不需要的消息。这些技巧之一是发送除用户母语之外的不同语言的垃圾邮件。根据 Message Labs 2009 年 7 月的情报报告 [1], 在法国、荷兰和德国, 垃圾邮件发送者使用垃圾邮件翻译技术生成垃圾邮件的比例分别为 53%、25% 和 46%。在中国和日本, 非英语垃圾邮件的比例甚至高达63.3%和54.7%。

机器对机器通信杂志, 卷。 1, 107–122。 doi:

10.13052/jmmc2246-137X.122

Qc 2014 河流出版社。版权所有。

该技巧相对有效，因为垃圾邮件发送者深入挖掘了当前垃圾邮件过滤机基于不擅长处理多语言电子邮件的词表检测垃圾邮件的缺陷。该报告 [1] 还解释说，通过充分利用自动翻译工具，垃圾邮件制造者制造了不同语言的垃圾邮件，并导致上述国家/地区的整体垃圾邮件增加了 13%。

自动化翻译工具的发展是自然而必要的。为了解决这个问题，应该提出垃圾邮件过滤机的多语言规则。在最近的一篇论文中，Quang-Anh Tran 等人。

[2]介绍了一种为SpamAssassin创建中文规则的方法。虽然这套规则做得很好，并被全球数以千计的电子邮件服务器共享，但它只能检测中文垃圾邮件。为了克服这种诡计，我们升级了方法并使其成为多语言的。换句话说，我们创建了一个系统，可以为不同的语言生成一套反垃圾邮件规则。实验是针对每种语言（中文、越南语和英语）和这三种语言的混合类型使用相同的数据集进行的。

本文的结构如下：在第二节中，我们回顾了一些针对特定语言的文件垃圾邮件的方法和相关知识。第三节介绍了我们方法的理论框架。接下来，进行实验，并在第 IV 节中比较结果。最后，第五节总结了本文并讨论了工作的未来。

2 相关作品

2.1 垃圾邮件刺客规则

SpamAssassin 是最常用的工具之一，用于确定电子邮件消息是垃圾邮件的可能性。它根据内容匹配规则过滤垃圾邮件。每条规则都有自己的分数。如果电子邮件获得足够的分数（超过预定义的阈值），它将被标记为垃圾邮件。

以下是 SpamAssassin 规则的示例：

图 1 是完整规则定义的示例。名为 FROM START WITH NUMS 的规则会根据正则表达式检查电子邮件的 FROM 标头是否以至少两个数字开头。如果电子邮件符合规则，它将为电子邮件的垃圾邮件分数添加一个分数。Schwartz (2004) [15] 详细描述了规则的剖析。为了有效地捕获特定语言的垃圾邮件，应该根据这些语言的特点生成规则。这就是我们旨在为国际环境建立多语言规则集的原因。

```

header FROM_STARTS_WITH_NUMS    From =~ /^d\d/
describe FROM_STARTS_WITH_NUMS  From: starts with nums
score FROM_STARTS_WITH_NUMS      0.390 1.574 1.044 0.579

```

图 1 SpamAssassin 规则示例

2.2 针对特定语言的SpamAssassin规则研究

Quang-Anh Tran 和他的合作伙伴 [2] 在他们的论文中解释说, 垃圾邮件检测分为两类: 基于规则和基于统计。第一个是指通过在电子邮件中搜索类似垃圾邮件的模式来执行检测。SpamAssassin 被称为基于规则的垃圾邮件检测机器中最受欢迎的代表。另一方面, 后者设法处理二分类问题; 垃圾邮件和火腿的数据集用于训练检测器。贝叶斯算法是使用最广泛的基于统计的垃圾邮件检测方法。Androutsopoulos (2000) [4] 和 Graham (2002) [5] 在这个主题上有典型的著作。此外, 还提出了其他基于统计的方法, 例如用于垃圾邮件检测的神经网络[6]、支持向量机[7]。

然而, 每种方法(基于规则或基于统计)都有其缺点。基于规则的方法很容易在服务器(或用户)之间共享, 但它是手动构建的。因此, 很难跟上垃圾邮件的快速变化。然而, 使用基于统计的方法, 只要训练数据集是最新的, 就更容易重新训练垃圾邮件检测器。但是, 不可能共享检测器的知识。因此, 他们提出了一种基于规则和基于统计的权衡混合方法来创建检测中文垃圾邮件的规则。该方法具有基于规则和基于统计的方法的所有优点: 检测器的快速训练和易于在服务器之间共享。

Nguyen T. A 等人。[8] 展示了一种基于语言分类检测越南垃圾邮件的方法。他们旨在引入越南语分段, 使用令牌选择来构建基于语言分类和贝叶斯组合的越南语垃圾邮件过滤器, 以充分支持越南语。比较了越南语分割和空间标记分割在垃圾邮件检测上的结果。他们的垃圾邮件检测率高出约 9%, 而垃圾邮件错误率降低了 3%。

尽管 [2] 和 [8] 中提出的两种方法都取得了积极的结果, 但它们只关注特定的语言。这里的问题是这些方法如何处理用户收到电子邮件的实际情况超过

每天一种语言，垃圾邮件发送者不断向世界各地的电子邮件用户发送多语言垃圾邮件。

3 理论框架

图 2 说明了我们的多语言规则是如何生成和维护的。

来自不同来源的电子邮件被分类并保存到 Spam & Ham 数据库中。分类由电子邮件用户和研究人员进行。因为这是一种基于内容的方法，所以我们对电子邮件所需要的只是主题和正文。在解码编码内容并剥离电子邮件附带的整个 html 标签后，我们使用 Google API 检测每封电子邮件的语言。对于每种语言（本文仅使用中文、越南语和英语三种语言），将调用合适的切分方法。这一步的产品是一个有意义的词表，它是下一步的输出。我们在下一部分讨论的其余过程中重用 [2] 中的算法。

多语言规则集通过三个步骤自动生成：模式检索、模式选择和分数分配。

3.1 模式检索

正如我们上面提到的，每种语言都有自己的方式将句子拆分成有意义的单词。对于某些语言，例如英语、法语或德语，可以通过空格轻松识别单词。然而，对于越南人来说，

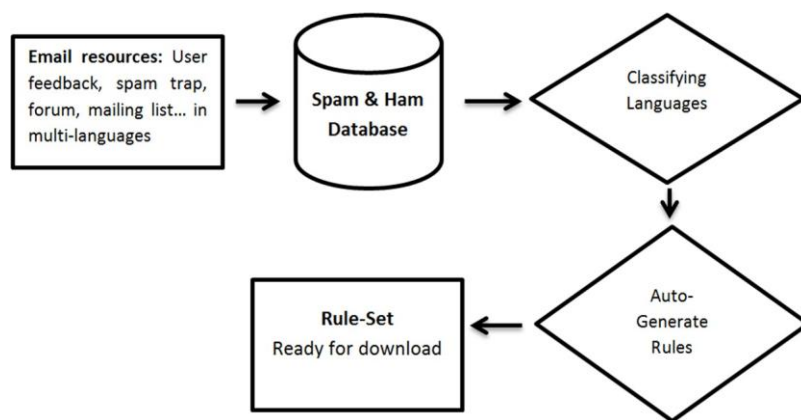


图2 多语言规则生成过程

中文或其他一些亚洲语言，这是不可能的，因为它们有特殊的语言单位，称为音节（越南语中的“Tie`n`g”或汉语中的“hanzi”）。

为了在处理电子邮件内容时达到最高效率，我们使用 Google Translate API [9] 来检测电子邮件的语言。尽管 API 运行良好且易于使用，但它不是免费的。然后，我们只在实验期间使用。为了进一步使用，我们考虑了一些其他解决方案，例如 [10] 中提供的 `Lingua::Identify` 和 `Guess-Language` [11]。在本文中，我们只实现了三种语言的分割：中文、越南语和英语。然而，我们正在努力提出一种多语言分词方法，正如 Guo-Wei Lee 在他的研究 [12] 中提到的那样。

对于中文电子邮件，我们完全应用了 [2] 中使用的中文分割技术，该技术基于以下方法：基于字典，最大匹配；从左到右。

对于越南语电子邮件，我们通过 Phuong Le-Hong [13] 提出的程序处理分词。该程序适用于越南语文本文件或文件夹，并将有意义的单词列表导出为 XML 格式。从这个 XML 文件中读取单词列表非常简单。

在英语电子邮件中拆分单词要简单得多，因为单词由空格分隔。我们刚刚找到并用换行符替换了空格符，并删除了句子中的所有标点符号。

3.2 模式选择

在按语言对电子邮件进行分类并提取有意义的单词后，我们应用一些模式选择方法分别为主题规则和正文规则选择好的模式。我们没有发现中文、越南语和英文电子邮件的选择模式有任何差异；因此，我们再次重用了 [2] 中的模式选择算法。

尽管基于 Yang [14] 的传统模式选择方法，该方法还是有一些变化。只有类似垃圾邮件的模式用于检测垃圾邮件。因此，修改了选择模式的公式。 V_{TS} 和 V_E 根据条件概率和贝叶斯定理计算如下：

$$P(H) = \frac{P(E \wedge H)}{P(E)} \quad (1)$$

$$电压 = P(E|H) = \frac{P(E \wedge H)}{P(H)} \quad (2)$$

其中:

- • V_{TS} 和 V_B 最能评估模式 t 与垃圾邮件、模式 t 与火腿之间的联系,即。
- • 选择具有最高比值 $R_t = V_{TS}/V_B$ 的前 N 个模式。 N 是规则集的大小。
- 小,是控制规则集性能的一个因素。
- E 是一条消息作为垃圾邮件出现的假设。
- H 是一条消息作为火腿出现的假设。

给定垃圾邮件和普通邮件数据集,对于模式 t , A 和 B 分别是垃圾邮件和普通邮件包含 t 的次数; C 和 D 分别是垃圾邮件和普通邮件不包含 t 的次数。(1)和(2)中的概率值计算如下。

$$P(E) = \frac{A+C}{A+B+C+D} \quad (3)$$

$$\overline{P(E)} = \frac{B}{A+B+C+D} \quad (4)$$

$$P(H) = \frac{A+B}{A+B+C+D} \quad (5)$$

$$P(E \wedge H) = \frac{A}{A+B+C+D} \quad (6)$$

$$\frac{P(E \wedge H)}{P(H)} = \frac{B}{A+B+C+D} \quad (7)$$

3.3 分数分配

这些规则是根据选定的一组类似垃圾邮件的模式创建的。有两种类型的规则:主体规则和主题规则。Henry Stern [15] 的 Fast SpamAssassin 分数学习工具用于为每个规则分配分数。

根据 Quang-Anh Tran 等人 [2] 的说明,实现了训练中性网络的“随机梯度下降”方法。该程序使用单个感知器和 (8) logsig 激活函数 (9) 将权重映射到 SpamAssassin 得分空间。

$$f(x) = \frac{1}{1 + e^{-\sum_{i=1}^N w_i x_i}} \quad (8)$$

$$y(x) = \frac{1}{1 + e^{-f(x)}} \quad (9)$$

其中 w_i 表示规则 i 的分数, x_i 描述给定消息是否激活规则 i , 传递函数 (8) 返回消息的分数。梯度下降用于训练神经网络。迭代调整网络参数以确保平均误差率始终下降。在不学习微积分的情况下, 具有线性传递函数、logsig 激活函数和均方误差函数的感知器的误差梯度如下:

$$E(x) = y(x) (1 - y(x)) (y_{\text{经验}} - y(x)) \quad (10)$$

使用以下函数更新权重:

$$w_i = w_i + \alpha E(x) x_i \quad (11)$$

其中, α 是学习率。该实现使用所谓的“随机梯度下降”方法, 该方法通过随机遍历训练集来进行增量更新, 而不是每个时期进行一批更新, 因为 SpamAssassin 规则命中是备用的。

4 实验

4.1 数据集

我们用于进行实验的数据分为 4 组。

电子邮件来自电子邮件用户的个人收件箱, 并被作者手动分类为垃圾邮件和非垃圾邮件。我们将所有电子邮件存储在 MySQL 数据库中。每种语言的垃圾邮件和非垃圾邮件都保存在具有相同结构 (ID (PK), 主题, 正文, 状态, 日期) 的单独表中, 然后, 有八个表为实验服务。

首先, 对三个第一组 (第 1、2 和 3 组) 进行实验, 以创建相应语言的规则。规则集仅基于单一语言数据集进行测试。结果保存为

表格1 数据集描述

团体	数垃圾邮件	数火腿的	语言
1	200	200	中国人
2	231	251	越南语
3	274	202	英语
4	705	653	多种语言

比较 (1) 。接下来，基于数据组 4 (多语言电子邮件) 测试单语言规则集以评估效率 (2) 。最后，基于数据组4生成混合语言规则集。记录该规则集的有效性并与结果 (1) 和 (2) 进行比较。

4.2 单语言规则集创建

具体的生成规则集的过程在第 3 节中提到。

对于中文规则，实验基于数据组 1，有 200 个 hams 和 200 个垃圾邮件。垃圾邮件检测率 (Spam Recall) 和失败警报率 (Ham Error) 如表2所示。

Chinese 规则给出了最好的结果，阈值等于 2.5，此时阳性真实率为 91.5%，并且消除了失败的警报率。

基于越南语进行的实验 (完全使用越南语数据集生成规则和测试规则——第 2 组) 也带来了积极的结果。

表 2 Chinese rule with Chinese dataset 的性能

临界点	垃圾邮件召回	火腿错误
0.5	93.5%	30.5%
1	91.5%	9.0%
1.5	91.5%	5.5%
2	91.5%	4.0%
2.5	91.5%	0.0%
3	91.5%	0.0%
3.5	85.0%	0.0%
4	75.5%	0.0%
4.5	71.0%	0.0%

表 3 Vietnamese rule with Vietnamese dataset 的性能

临界点	垃圾邮件召回	火腿错误
0.5	90.5%	34.7%
1	87.4%	27.9%
1.5	83.1%	11.2%
2	81.4%	2.4%
2.5	81.4%	0.0%
3	78.4%	0.0%
3.5	73.6%	0.0%
4	66.2%	0.0%
4.5	59.3%	0.0%

表 4 英语规则与英语数据集的性能

临界点	垃圾邮件召回	火腿错误
0.5	98.5%	81.2%
1	97.1%	50.5%
1.5	96.0%	24.3%
2	95.6%	5.0%
2.5	95.3%	0.0%
3	93.1%	0.0%
3.5	87.6%	0.0%
4	82.8%	0.0%
4.5	60.2%	0.0%

在阈值 0.5 时，垃圾邮件召回率确实很高（90.5%），但非正常邮件错误率是不可接受的（高达 34.7%）。但是，当我们提高阈值时，结果会越来越好。特别是，在阈值 2（2.4%）处，ham 错误率显著下降，而垃圾邮件召回率与之前的阈值（81.4%）相比保持不变。

我们做了同样的事情来生成英文规则集，然后在仅基于英文电子邮件测试规则后记录结果。结果显示在表 4 中。

英语规则集在以下门槛上非常有效

2.5. 在这一点上，正确率保持在 95% 以上，而 ham 错误被完全消除。

在完成生成 SpamAssassin 规则并使用相应语言测试这些规则集的实验后，我们在真实垃圾邮件检测率和警报失败率方面取得了非常积极的结果。然而，这些规则集是否仍然适用于多语言数据集？答案将在下一个实验中揭晓。

4.3 使用多语言电子邮件测试的单语言规则集

为了评估在多语言环境中的效率，使用包含多语言705个垃圾邮件和653个垃圾邮件的数据组4测试了中文、越南语和英语的三组规则。表 5 显示了中文、越南语和英语规则集如何工作的结果。

统计数据清楚地表明，在使用多语言电子邮件数据集时，所有规则集在真实阳性率方面的表现都非常差，尤其是中文规则，在阈值 0.5 时仅检测到 24.5%，而上次实验则超过 93%。English rule 和 Vietnamese 更擅长垃圾邮件检测，但结果远低于 working 时的结果

表 5 单语言规则与多语言数据集中文的性能

临界点	英语				越南语	
	垃圾邮件	火腿	垃圾邮件	火腿	垃圾邮件	火腿
	记起	错误	记起	错误	记起	错误
0.5	24.5%	0.2%	60.6%	3.5%	51.9%	4.9%
1	21.8%	0.2%	59.0%	2.0%	49.9%	1.8%
1.5	20.7%	0.0%	54.8%	0.2%	44.3%	1.1%
2	19.0%	0.0%	53.6%	0.2%	42.4%	0.3%
2.5	18.2%	0.0%	49.4%	0.0%	41.6%	0.0%
3	16.9%	0.0%	46.8%	0.0%	40.7%	0.0%
3.5	16.7%	0.0%	38.4%	0.0%	39.9%	0.0%
4	16.0%	0.0%	20.7%	0.0%	32.5%	0.0%
4.5	15.6%	0.0%	14.8%	0.0%	22.1%	0.0%

具有单一语言数据集。对坠落的解释非常清楚和直接。从特定语言生成的规则只能有效地检测该语言的垃圾邮件。因此，我们期待一个多语言规则集，可以在大量的多语言电子邮件中有效地检测垃圾邮件。

4.4 多语言规则集

最后的实验是从包含中文、越南语和英文三种语言的电子邮件的数据组 4 中生成并测试规则集。

在对每封邮件的语言进行分类后，我们对同一语言的邮件集合进行了分词。模式选择选择最佳模式来评估模式是否喜欢垃圾邮件。根据选定的模式，规则集会自动生成。Fast SpamAssassin 分数学习工具将通过为每个规则分配分数来处理其余部分。将这些步骤应用于多语言数据集，我们获得了一组用于检测垃圾邮件的多语言规则。

表 6 说明了基于多语言规则集检测垃圾邮件的测试结果。在阈值的第一级，虽然垃圾邮件召回率最高（94.17%），但非正规邮件错误率高达 48.80%。它不能容忍一组 SpamAssassin 规则。然而，当阈值增加到 2.5 时，结果要好得多。真阳性率为89.40%，假阳性率为0%。在此阈值下，对于相同的多语言数据集，中文规则、越南规则和英文规则的性能分别为 18.2%、49.4% 和 41.6%。这一比较证明，从多语言数据集生成的规则比仅基于单一语言生成的规则更有效。

表 6 多语言规则与多语言数据集的性能

临界点	垃圾邮件召回	火腿错误
0.5	94.17%	48.80%
1	92.00%	29.13%
1.5	90.20%	13.67%
2	89.50%	3.80%
2.5	89.40%	0.00%
3	87.67%	0.00%
3.5	82.07%	0.00%
4	74.83%	0.00%
4.5	63.50%	0.00%

5 评论

通过以上的实验，生成了三组单语言规则。表 3、4 和 5 中的结果表明，这些规则集在处理单语言数据集（大多数电子邮件只有一种语言）时有效。在同样的阈值 2.5 下，中文规则可以检测到高达 91.5% 的垃圾邮件，越南规则可以检测到 81.4% 的垃圾邮件，英文规则的结果为 95.3%，失败警报为 0%。然而，当将这些单语言规则集应用于检测多语言垃圾邮件时，结果并不理想。具体而言，在阈值 2.5 下，中文、越南语和英文规则的垃圾邮件检测百分比分别为 18.2%、49.4% 和 41.6%。这种下降的原因很清楚。每组单语言规则都是根据相应的语言数据集生成的。这意味着该规则只能处理该语言的垃圾邮件。因此，考虑并评估了一组多语言规则。这套规则建立

来自包括汉语、越南语和英语在内的多语言数据集。

运行一个实验来评估多语言规则集的性能。结果是积极和有希望的。在阈值为 2.5 时，垃圾邮件检测率为 89.40%，同时消除了非正常错误。从这些发现中可以看出，在国际工作环境中应用时，多语言规则集在检测垃圾邮件方面的有效性。

6 结论

生成基于特定语言的垃圾邮件检测规则是对抗垃圾邮件发送者的正确方法。然而，为了处理一个电子邮件服务器接收多种语言的电子邮件，我们需要一个扩展的解决方案。因此，我们升级了[2]中提出的方法来实现

能够基于多语言数据集自动生成多语言规则的系统。实验结果表明，与基于单一语言数据集生成的规则相比，这些规则有助于 SpamAssassin 更准确地检测垃圾邮件。

尽管取得了积极成果，但在未来我们还需要解决一些问题。首先，应该分析一种检测电子邮件语言的新方法。目前的成本如此之高。其次，我们期待有更好的算法来检索原始电子邮件的模式。最后，如果系统面临大量语言，由于缺乏[12]中提到的通用分词方法，分词将是一个大问题。

致谢

这项研究得到了越南国家科学技术发展基金会 (NAFOSTED) 的支持，项目编号为 102.01-2010.09。这项工作得到了新南威尔士大学 (UNSW) Vice 研究计划的大力支持。

参考文献

- [1] 多语言垃圾邮件 - 垃圾邮件发送者的新趋势。可用的：
<http://www.spamfighter.com/News-12908-Multi-language-Spam-垃圾邮件发送者之间的新趋势.htm>
- [2] Tran, Q. A., Duan, H. X. Li, X., “垃圾邮件检测的实时统计规则”IJCSNS 国际计算机科学与网络安全杂志, 第 6 卷第 2B 期, 第 178-184 页, 2006 年 2 月。
- [3] Androutsopoulos I.、Koutsias, J.、Chandrinou, K.V.、Paliouras, G.、Spyropoulos, C.D., “朴素贝叶斯反垃圾邮件过滤器的评估”, 新信息时代机器学习研讨会论文集, 第 9 页-17, 第 11 届欧洲机器学习会议, 西班牙巴塞罗那, 2000 年。
- [4] Graham, P., “针对垃圾邮件的计划。网络文档”(2002 年)。可用的: <http://www.paulgraham.com/spam.html>.
- [5] Drucker, H., Wu, D., Vapnik V., “用于垃圾邮件分类的支持向量机”, IEEE Transaction on Neural Networks.10(5), 1048-1054, 1999。

- [6] Özgür, L., Güngör, T., Gürgen, F., “粘着语言的自适应反垃圾邮件过滤：土耳其语的特例”, Pattern Recognition Letters, 1819–1831 25. 2004 年。
- [7] Nguyen T. A.、Tran Q. A.、Nguyen N. B., “基于语言分类的越南垃圾邮件检测”, HUT-ICCE 2008 – 第二届通信与电子国际会议, 越南会安, 2008 年。
- [8] 谷歌翻译 API。可用: <https://developers.google.com/translate/>
- [9] 语言::识别。可用的: <http://search.cpan.org/ambs/Lingua-Identify-0.51/lib/Lingua/Identify.pm>
- [10] 猜语言。可在<http://code.google.com/p/guess-language/>
- [11] Guo-Wei Lee, “基于决策树和集成特征选择算法的多语言垃圾邮件过滤机制”, 1997 年。
- [12] Phuong Le-Hong 等人, “越南文本分词的混合方法”, 第二届语言和自动机理论与应用国际会议论文集, LATA, Springer LNCS 5196, 西班牙塔拉戈纳, 2008 年。
- [13] Yang, Y., Pedersen, J. O., “文本分类中特征选择的比较研究”, 第 14 届国际机器学习会议论文集, 第 412–420 页, 1997 年。
- [14] Fast SpamAssassin 分数学习工具。可用的: <http://spamassassin.apache.org>
- [15] Schwartz, “SpamAssassin”, 2004 年, O’Reilly。

传记



Vu Minh Tuan 是河内大学信息技术学院的讲师和研究协调员。目前, 他还在 IP Communications, JSC 的软件部门担任项目经理。

2012 年，他开始在英国中央兰开夏大学攻读理学硕士学位。2006年至2010年就读于河内大学，一直从事SpamAssassin的越南规则研究，以及反垃圾邮件、电子邮件系统和数据挖掘领域的一些项目。



Tran Quang Anh 是河内大学信息技术副教授。分别于2003年和2001年在清华大学获得博士和硕士学位。1997年毕业于华中科技大学，主要研究方向为网络安全进化算法和现场可编程门阵列。



蒋博士获得理学学士学位。系统与控制工程学士学位和硕士学位分别于1997年和1999年在中国和澳大利亚获得计算机科学工程（研究）学位。2008 年，他成功获得澳大利亚研究生奖（APA）奖学金，在悉尼科技大学（UTS）完成了通信工程和软件工程博士学位。在加入 UTS 之前，姜博士

1999年至2003年在澳大利亚悉尼从事VoIP行业5年软硬件工作经验，博士毕业后在工程与信息技术学院（FEIT）担任研究助理，后任研究员3年），悉尼科技大学。此外，他在 UTS 拥有 5 年的讲师教学经验和 2 年的学科协调员经验。目前，他在新南威尔士大学（UNSW）担任讲师和全职 UNSW 副校长任命的研究员。他在计算智能及其应用领域发表了 60 多篇国际期刊和会议论文。他目前的研究兴趣包括数据分析、仿生算法和元启发式算法、水下通信、网络安全、自主通信网络、智能和移动代理、网络协议。



Van Quan, Tran Quan 于 2012 年毕业于河内大学（HANU），获得理学学士学位。信息技术学士学位，目前（2014 年）是英国中央兰开夏大学的信息系统设计硕士生。他有在 HANU 担任助教的教学经验和超过 2 年的开发人员工作经验。他目前的研究兴趣包括：文本挖掘、语音识别和人机交互。

