



## 星环利用大数据分析提升垃圾短信过滤效果

### 客户背景

#### 通信发展伴生的弊端

移动通信技术的不断发展之下，短信已经成为人们生活中不可或缺的工具之一，但伴随而来的垃圾短信泛滥，不但占用了电信运营商宝贵的网络资源，而且给人们的生活、工作带来了无尽的烦恼。如何对垃圾短信进行智能识别与实时监测，从而提高客户满意度与服务质量，成为了当前电信行业亟待解决的问题。

#### 不断加剧的垃圾短信问题

在通讯信息技术不断发达的当今社会，短信作为一种直达用户，成本低廉的方式，不但被越来越多的商业公司用于促销或宣传目的，甚至成为不法分子实施诈骗的重要手段。中国信息产业部报告显示，2014年，全国移动短信业务量7630.5亿条，而垃圾短信的数量就占了1/4左右，这些造成了对用户的骚扰甚至财产损失的垃圾短信问题正在变得越来越严重。

### 面临挑战

**增强垃圾短信检测手段：**挖掘垃圾短信的隐藏信息，利用更新的技术手段，提升垃圾短信防范效率。

**提高垃圾短信检测精度：**传统单纯以字符串匹配过滤垃圾短信的方法误检率较高，而且事后增加关键词的手段存在滞后性。

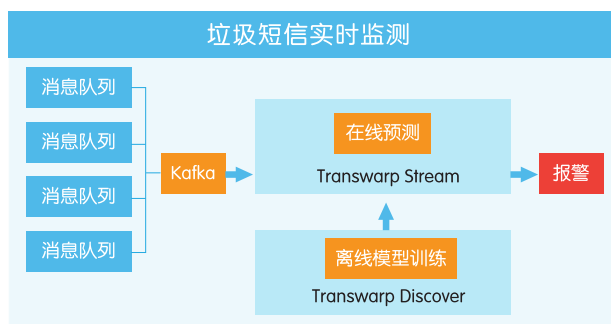
**实时监测：**对短信实时监测并完成垃圾短信的过滤，不断降低垃圾短信到达率，提高用户满意度。

### 解决方案

**部署垃圾短信实时监测平台：**基于分布式消息队列Kafka和流处理引擎实现实时的垃圾短信判断和预警。同时，结合人工确

垃圾短信数据，加入训练集用于机器自动学习，垃圾短信判断准确率99.9%以上。

垃圾短信实时监测平台，实现垃圾短信的实时过滤：针对短信数据24小时不间断产生、大规模、高并发等特点，星环科技基于Transwarp Stream流处理计算框架研发垃圾短信实时监测平台，短信数据通过实时消息队列进入计算集群，利用事先训练好的模型完成数据转换、特征提取、分析及实时判断预警等复杂计算。模型判断出的垃圾短信会通过人工确认，人工判断确实为垃圾短信的数据会加入训练集用于模型的迭代训练。



星环科技垃圾短信实时监测平台流程图

**流计算引擎实现实时垃圾短信过滤：**测试员模拟典型的短信发送场景，对10万条短信数据集连续测试两小时。测试结果显示，每服务器节点每秒钟可对1000-3000条短信实施过滤计算，正常短信和垃圾短信的识别率均在99.9%以上。

**星环科技工程师程大伟说：**“在垃圾短信实时监测平台中，网络适配器不但是短信数据传输至流处理引擎的通道，而且是影响整个平台性能的瓶颈，需要在90%以上高负荷的情况下保持数据传输的稳定性。”

部署垃圾短信实时监测平台后，垃圾短信在到达用户之前即被实时过滤，而机器学习的机制可以不断根据短信内容调整或更新训练模型。“显然，垃圾短信到达率的降低将帮助运营商提高用户的满意度。此外，运营商可以根据短信规模灵活调整计算集群的服务器数量满足过滤需求，更重要的是，运营商原来被垃圾短信占用的网络资源被释放，运营商可借此改进业务运营。”程大伟补充到。

## 实施效果

**提高用户满意度：**99.9%以上的垃圾短信被自动过滤，通信公司的用户对垃圾短信的抱怨和投诉率显著降低，提高其用户满意度。

**提升网络资源利用率：**运营商将垃圾短信占用的网络资源用于其它短信业务，改进短信业务市场运营。

**实时垃圾短信监控：**通过对垃圾短信的实时监控和持续的机器自动学习手段，不断满足越来越复杂的垃圾短信管理和过滤需求。

**提高垃圾短信过滤效率和精度：**电信运营商根据垃圾短信的内容整理出特定的关键词，在短信到达用户前，利用字符串匹配的方式在数据中心实施计算和过滤。

“这种基于字符串匹配的手段存在明显的不足。首先是滞后性，运营商只能在事后整理关键词，然而发送垃圾短信的用户会不断的测试关键词并调整短信内容，例如：加入特殊符号、利用同音汉字等等，这使得依靠整理和分析关键词的方式实现垃圾短信过滤效果越来越差。”程大伟说：“其次，伴随关键词的数量不断增加，服务器的计算性能对垃圾短信过滤效率影响越来越大。”

此外，对于电信运营商而言，这些垃圾短信的传送占用和浪费了大量宝贵的网络资源。

电信运营商急需利用更新的技术手段对海量短信数据实现实时运算和分析，挖掘出有价值的垃圾短信参考依据，形成机器学习的机制，实现自动化的垃圾短信过滤，从而充分利用网络资源，并提高用户满意度。

星环科技流处理引擎提供强大的流计算表达能力，支持在流数据上进行实时事件检测和批处理、机器学习等复杂的计算逻辑，是电信运营商提高垃圾短信防范的理想工具。

## 目标与展望

下一步，星环科技将和电信运营商深入合作，继续挖掘短信发送用户和接受用户之间的关系，利用图的方法进一步提升垃圾短信过滤效率和质量。

星环信息科技有限公司（上海）有限公司

📍 地址：上海市徐汇区桂平路481号18幢3层301室（漕河泾新兴技术开发区）

✉ 邮编：200233 ☎ 电话：4008 079 976

🌐 网址：www.transwarp.io

**TRANSWARP**  
DATA HUB