ОПТИМИЗАЦИЯ АНТИВИРУСНОЙ ЗАЩИТЫ ЧЕРЕЗ МАШИННОЕ ОБУЧЕНИЕ: АНАЛИЗ ПРЕИМУЩЕСТВ И ВЫЗОВОВ

Бойков Илья Александрович

студент, Московский государственный технический университет им. Н. Э. Баумана, РФ, г. Москва

Цыганок Дарья Игоревна

студент, Московский государственный технический университет им. Н. Э. Баумана, РФ, г. Москва

В последние десятилетия, кибербезопасность столкнулась с беспрецедентными вызовами в связи с постоянно растущим количеством и сложностью киберугроз. В этой динамичной обстановке машинное обучение (ML) оказалось ключевым инструментом в борьбе с вирусами и другими видами вредоносного ПО. Ранее антивирусные программы в основном опирались на сигнатурный подход, который предполагал обнаружение и блокирование угроз на основе уже известных вирусных сигнатур. Однако этот метод быстро устаревает, поскольку новые вирусы постоянно разрабатываются и модифицируются, делая их невидимыми для традиционных сигнатурных сканеров.

Внедрение машинного обучения радикально изменило эту ситуацию. В отличие от традиционных методов, ML позволяет системам антивирусной защиты обучаться и адаптироваться к новым угрозам в реальном времени, повышая их способность проактивно распознавать и блокировать вредоносные программы, даже если они ранее не встречались.

Примеры алгоритмов машинного обучения, используемых в антивирусной защите

В сфере антивирусной защиты, машинное обучение представлено разнообразными алгоритмами, каждый из которых выполняет уникальную роль. Нейронные сети, в том числе глубокие нейронные сети, стоят в авангарде обнаружения новых и неизвестных угроз. Они обучаются на обширных наборах данных, содержащих информацию о безопасных и вредоносных программах, что позволяет им эффективно распознавать и идентифицировать угрозы. Примеры таких систем включают Convolutional Neural Networks, которые анализируют структуру вредоносного ПО и выявляют аномалии в коде или поведении приложений.

Алгоритмы кластеризации, такие как K-средних, играют важную роль в группировке похожих видов угроз. Это позволяет выявлять новые варианты уже известных вирусов на основе общих характеристик, даже если эти вирусы были модифицированы или зашифрованы. Деревья решений и случайные леса, являясь ансамблями деревьев решений, улучшают стабильность и точность классификации вредоносных программ по различным атрибутам и поведенческим характеристикам. Они особенно эффективны в выявлении фишинговых сайтов или опасных файлов на основе их специфических особенностей.

Метод опорных векторов (SVM) представляет собой мощный алгоритм классификации, который разделяет данные на классы с помощью гиперплоскости в многомерном пространстве. В антивирусной защите SVM используются для различения вредоносных и безопасных программ на основе анализа их характеристик и поведения. Наконец, алгоритмы обнаружения аномалий, такие как Изолирующий лес, специализируются на выявлении аномальных данных или поведения, что может указывать на новую или неизвестную

вредоносную активность. Они особенно полезны для обнаружения необычных сетевых паттернов или изменений в системных файлах, которые могут быть признаками вредоносной деятельности.

Алгоритмы обнаружения аномалий, такие как Изолирующий лес, играют ключевую роль в выявлении необычных или ненормальных наблюдений в данных. Они специально разработаны для обнаружения тех случаев, которые отклоняются от установленной нормы, что делает их особенно ценными для идентификации новых или неизвестных вредоносных программ. В контексте антивирусной защиты, эти алгоритмы могут обнаруживать необычные поведенческие паттерны или активности в системе, которые могут сигнализировать о наличии вредоносного ПО. Например, они могут выявлять аномальное сетевое поведение или неожиданные изменения в системных файлах, которые часто являются признаками вредоносной деятельности.

Антивирусные решения с использованием новых технологий представляют собой важный шаг в борьбе с киберугрозами, преодолевая ограничения традиционных методов и предоставляя улучшенные возможности для обнаружения и реагирования на угрозы.

Новые методы помогают обнаруживать новые угрозы без опоры на сигнатуры, анализируя и сравнивая поведение программ с известными вредоносными образцами. Это особенно важно в условиях постоянно изменяющихся и адаптирующихся вирусов и вредоносных программ. [1]

Эти технологии обеспечивают способность адаптироваться к новым и ранее не виданным угрозам. Системы могут самостоятельно 'учиться' на основе новых данных, что делает их способными предсказывать и противостоять новым видах атак. [2]

Интеграция передовых технологий значительно автоматизирует процесс обнаружения и реагирования на угрозы. Это включает в себя автоматическое обновление баз данных вирусов и сканирование на наличие угроз.

Многие компании в области кибербезопасности активно интегрируют передовые технологии в свои продукты, улучшая их способность обнаруживать и противостоять различным видам угроз.

Машинное обучение в антивирусной защите несет в себе определенные риски и ограничения, которые важно учитывать. Одна из основных проблем заключается в возможности ошибок обнаружения, включая ложные срабатывания и упущенные угрозы. Эти ошибки возникают из-за того, что алгоритмы могут не всегда точно распознавать вредоносные программы, что приводит к неправильной классификации безопасных программ как вредоносных или наоборот.

Еще одним важным недостатком является зависимость машинного обучения от качества и объема обучающих данных. Если данные для обучения модели недостаточны или некорректно размечены, это может снизить точность и эффективность системы. Неправильная разметка данных может привести к тому, что алгоритмы будут делать неверные выводы, и эти ошибки будут накапливаться со временем, ухудшая общую производительность системы.

Также следует учитывать высокую требовательность машинного обучения к вычислительным ресурсам и сложности его интеграции с существующими системами безопасности. Это может стать вызовом, особенно в условиях ограниченных ресурсов или там, где требуется совместимость с уже существующими технологиями.

Кроме того, применение машинного обучения в антивирусной защите может представлять угрозы для конфиденциальности и безопасности данных. При использовании больших объемов данных для обучения моделей возникает риск их утечки или неправомерного использования, что может стать серьезной проблемой с точки зрения защиты личной информации и коммерческих данных.

Один из заметных примеров использования машинного обучения в антивирусной защите демонстрирует Microsoft с их подходом к борьбе с вариантом вредоносного ПО GoldMax.

Используя сочетание "fuzzy hashing" и глубокого обучения, основанного на техниках обработки естественного языка и компьютерного зрения, Microsoft успешно блокировала эту угрозу. Данный метод позволил улучшить точность классификации вредоносных программ и ускорить их обнаружение, что критически важно для защиты от сложных кибератак. [3]

Применение машинного обучения в антивирусных решениях демонстрирует значительные успехи в обнаружении и предотвращении киберугроз, преодолевая ограничения традиционных подходов. Адаптивность и обучаемость этих систем позволяют им эффективно противостоять новым и адаптирующимся вирусам, значительно улучшая защиту. Однако важно учитывать риски, связанные с возможными ошибками обнаружения, зависимостью от качества обучающих данных, высокими требованиями к вычислительным ресурсам, а также проблемами конфиденциальности и безопасности данных. Тщательный подход к интеграции машинного обучения и постоянное улучшение алгоритмов остаются ключевыми для достижения оптимальной эффективности антивирусных систем.

Список литературы:

1. The State of Ransomware in the U.S.: Report and Statistics 2023 // Emsisof URL: https://www.emsisoft.com/en/blog/35668/the-pros-cons-and-limitations-of-ai-and-machine-learning-in-antivirus-software/] (дата обращения: 18.12.2023).

2. What is the role of machine learning in antivirus? // James Parker URL: https://www.jamesparker.dev/what-is-the-role-of-machine-learning-in-antivirus/] (дата обращения: 18.12.2023).

3. Combing through the fuzz using fuzzy hashing and deep learning to counter malware detection evasion techniques // Microsoft URL: https://www.microsoft.com/en-us/security/blog/2021/07/27/combing-through-the-fuzz-using-fuzzy-hashing-and-deep-learning-to-counter-malware-detection-evasion-techniques/] (дата обращения: 18.12.2023).