

АНАЛИЗ МЕТОДОВ ПРОГНОЗИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ

Запорожец Д.Б., Скулиш М.А.

Институт телекоммуникационных систем

КПИ им. Игоря Сикорского, Украина

E-mail: zaporozhetsdb@gmail.com

Analysis of Forecasting Methods using Neural Networks

Was analyzed forecast of Internet traffic as an important issue that has received few attentions from the computer networks field. By improving this task, efficient traffic engineering and anomaly detection tools can be created, resulting in economic gains from better resource management. This paper presents a Neural Network Ensemble (NNE) for the prediction of TCP/IP traffic using a Time Series Forecasting (TSF) point of view. Several experiments were devised by considering real-world data from two large Internet Service Providers. In addition, different time scales (e.g. every five minutes and hourly) and forecasting horizons were analyzed. Overall, the NNE approach is competitive when compared with other TSF methods (e.g. Holt-Winters and ARIMA).

Прогнозирование трафика TCP / IP является важной задачей для любого среднего или крупного сетевого провайдера [1]. Несколько провайдеров интернет-услуг (ISP) используют Multiprotocol Label Switching (MPLS) для реализации инфраструктуры трафика путем создания полной сетки виртуальных сетей MPLS между всеми парами маршрутизаторов в сети [2]. Теоретически это позволит оптимизировать ресурсы полосы пропускания в сети, что позволит повысить качество обслуживания в целом. В случаях, когда MPLS не используется, традиционное распределение весов протокола маршрутизации может быть сделано гораздо более эффективно, зная будущую матрицу трафика спроса между всеми конечными точками сети. Следовательно, нужно предусмотреть автоматические средства разработки трафика, которые адаптируются к будущим условиям сети на основе точных алгоритмов прогнозирования трафика.

Прогнозирование трафика также может помочь обнаружить аномалии в сетях [3]. Атаки безопасности, такие как отказ в обслуживании, вирусы или даже нерегулярное количество СПАМ, могут быть обнаружены путем сравнения реального трафика со значениями, предсказанными алгоритмами прогнозирования.

В настоящее время эта задача часто решается интуитивно опытными сетевыми администраторами с помощью маркетинговой информации о будущем количестве клиентов и их обычном поведении [1]. Но это не дает точную картину о том, как будет выглядеть трафик, мало используемого для серьезного ежедневного сетевого администрирования. С другой стороны, вклад из областей оперативных исследований, статистики и аналитики привели к

более точным методам прогнозирования, которые заменили основанные на интуиции. В частности, поле прогнозирования временных рядов (TSF), также называемое одномерным прогнозированием, имеет дело с предсказанием хронологически упорядоченной переменной [5], [6]. Цель TSF - моделировать сложную систему как «черный ящик», предсказывая ее поведение, основанное на исторических данных, а не как это работает.

Из-за его важности было предложено несколько методов TSF, таких как Holt-Winters [7], методология ARIMA [8] и нейронные сети [9]. Holt-Winters была разработана для серий с трендовыми и сезонными факторами. Совсем недавно была предложена двойная сезонная версия. ARIMA является более сложным подходом, требующим таких шагов, как идентификация модели, оценка и валидация. Каждая модель ARIMA основана на линейной комбинации прошлых значений и / или ошибок. Нейронные сети - это модели подключений, вдохновленные поведением центральной нервной системы, и, в отличие от предыдущих методов, они могут прогнозировать нелинейные ряды. В прошлом несколько исследований доказали предсказуемость сетевого трафика с использованием аналогичных методов, таких как HoltWinters и ARIMA [3]. Следуя свидетельствам нелинейного сетевого трафика, также были предложены NN [4].

Данный подход использует уже имеющуюся информацию, предоставленную Simple Network Management Protocol (SNMP), которая с достаточной точностью оценивает трафик, проходящий через каждый сетевой интерфейс. SNMP широко используется каждым интернет-провайдером, поэтому сбор этих данных не вызывает дополнительного трафика в сети.

Naive Benchmark Method. Наиболее часто используемым методом наивного прогнозирования является прогнозирование будущего. Тем не менее, этот метод будет плохо работать в сезонных данных. Таким образом, лучшей альтернативой является использование сезонной версии, где прогноз будет определяться наблюдаемым значением за тот же период, связанный с предыдущим сезонным циклом.

B. Holt-Winters. Важный метод прогнозирования из семейства методов экспоненциального сглаживания. Прогностическая модель основана на некоторых основных шаблонах (например, трендовых и сезонных), которые отличаются от случайных шумов путем усреднения исторических значений [7]. Его популярность обусловлена такими преимуществами, как простота использования, сокращенный спрос на вычислительные ресурсы и точность прогнозов, особенно при совместном использовании с сезонными сериями.

C. ARIMA Methodology. Основное преимущество этого метода зависит от точности над более широкой областью рядов. Глобальная модель основана на линейной комбинации прошлых значений и ошибок, называемой AutoRegressive Integrated Moving-Average (ARIMA).

D. Artificial Neural Networks. Нейронные модели являются врожденными кандидатами на прогнозирование из-за их нелинейных и помехоустойчивых возможностей. Основная идея – обучить нейронные сети прошлым данным, а затем использовать эту сеть для прогнозирования будущих значений.

В этой работе было рассмотрено нескольких подходов к прогнозированию при применении к трафику TCP / IP. К ним относятся три метода прогнозирования временных рядов (TSF): Holt-Winters, методология ARIMA и подход нейронной сети (NNE).

Анализ преимуществ и недостатков приведенных алгоритмов показал, что следует методы прогнозирования с использованием нейронных сетей к активным сценариям измерения, в которых информация уровня пакета в реальном времени подается в механизм прогнозирования или рекомендуем применять методы прогнозирования для требований к трафику, связанных с конкретными интернет-приложениями, поскольку это может принести пользу некоторым управленческим операциям, выполняемым интернет-провайдерами, таким как приоритизация трафика и распределение сетевых ресурсов. Кроме того, для улучшения этапа выбора модели в тестируемых моделях TSF можно использовать ряд методов оптимизации. В этом случае оптимизация модели может проводиться параллельно с использованием модели прогнозирования в режиме реального времени, выполняя замену модели, когда ожидаются лучшие результаты.

Література

1. K. Papagiannaki, N. Taft, Z. Zhang, and C. Diot, “Long-Term Forecasting of Internet Backbone Traffic,” *IEEE Transactions on Neural Networks*, vol. 16, no. 5, pp. 1110–1124, Sept. 2005.
2. B. Davie and Y. Rekhter, *MPLS: Multiprotocol Label Switching Technology and Applications*. USA: Morgan Kaufmann, 2000
3. B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen, “Sketch-based Change Detection: Methods, Evaluation, and Applications,” in *Proceedings of Internet Measurement Conference (IMC’03)*. Miami, USA: ACM, Oct. 2003.
4. J. Jiang and S. Papavassiliou, “Detecting Network Attacks in the Internet via Statistical Network Traffic Normality Prediction,” *Journal of Network and Systems Management*, vol. 12, pp. 51–72, 2004.
5. J. Hanke and A. Reitsch, *Business Forecasting*. Allyn and Bacon Publishing, Massachusetts, USA, 1989.
6. S. Makridakis, S. Wheelwright, and R. Hyndman, *Forecasting: Methods and Applications*. John Wiley & Sons, New York, USA, 1998.
7. P. R. Winters, “Forecasting sales by exponentially weighted moving averages,” *Management Science*, vol. 6, pp. 324–342, 1960.
8. G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco, USA, 1976.
9. A. Lapedes and R. Farber, “Non-Linear Signal Processing Using Neural Networks: Prediction and System Modelling,” Los Alamos National Laboratory, USA, Technical Report LA-UR-87-2662, 1987.