

Ю.А. Скудняков, В.А. Мороз, Н.Н. Гурский
**ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ AZURE MACHINE LEARNING
В РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМАХ**

Белорусский государственный университет информатики и радиоэлектроники, г. Минск
Белорусский национальный технический университет, г. Минск

Рассмотрены основные возможности среды AzureMachineLearning для создания модели с целью ее дальнейшего использования в разрабатываемой рекомендательной системе. Построен тестовый эксперимент, в котором получена обученная модель для выдачи рекомендаций по фильмам для заданных пользователей.

Ключевые слова: рекомендательные системы; машинное обучение; AzureMachineLearning; ранжирование; релевантность.

В настоящее время неотъемлемой частью большинства современных интернет-приложений являются рекомендательные системы, которые позволяют на основании наборов атрибутов профилей пользователей и исторических сведений об их активности в системе выделить группы интересов клиентов, а также сформировать множество предположительно интересных для конкретных потребителей элементов, которые затем могут быть рекомендованы им.

Большое количество таких систем основано на технологиях машинного обучения. Машинное обучение - это метод обработки и анализа данных, который позволяет выполнить обучение компьютеров на основе имеющихся данных с целью прогнозирования будущего поведения, результатов и тенденций. Технологии машинного обучения являются одними из наиболее перспективных подходов к анализу больших данных, поскольку благодаря этим методам, компьютер может находить в огромных массивах данных изначально неизвестные закономерности и взаимосвязи. Подобные алгоритмы могут быть широко использованы в различных приложениях для рекомендации, например, музыки, фильмов, книг, основываясь на накопленных знаниях о предпочтениях пользователей [1].

Существует различные инструменты для использования машинного обучения, которые осуществляют интеллектуальный анализ данных, например: Azure Machine Learning, Amazon Machine Learning, BigML, Scikit-learn, Apache Mahout, PyLearn2 and NuPicидр.

Одним из наиболее мощных является Azure Machine Learning - облачный сервис для выполнения задач прогнозной аналитики (predictive analytics). Машинное обучение Azure включает сотни встроенных пакетов и поддержку настраиваемого кода. Данный сервис позволяет осуществлять построение и использование сложных моделей машинного обучения в достаточно простой и наглядной форме. Используя Azure Machine Learning, можно легко опубликовать свои алгоритмы в виде отдельных сервисов, которые в последующем могут быть использованы в пользовательском приложении.

В Azure Machine Learning доступны следующие группы алгоритмов [2].

- алгоритмы обнаружения аномалий;
- алгоритмы классификации;
- алгоритмы кластеризации;
- алгоритмы регрессии.

Целью данной работы является рассмотрение возможностей Azure Machine Learningstudio, которые могут быть применены для создания рекомендательной системы. Для рассмотрения выбран набор данных, содержащий информацию об оценках пользователей по просмотренным ими фильмам, и на основе этих данных построена модель, которая затем может

СЕКЦИЯ 5.1 ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА (СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ)

быть использована для получения рекомендаций по фильмам, которые возможно будут интересны для заданного пользователя.

В качестве рассматриваемых наборов данных были взяты предоставленные для изучения в самой среде AzureMachineLearning тестовые наборы данных – Movieratings и IMDB MovieTitles. Первый набор данных содержит 227472 строк, каждая строка включает следующую информацию: уникальный идентификатор пользователя (UserId), уникальный идентификатор фильма (MovieId), оценку (Rating) и метку времени (Timestamp). Фрагмент содержимого набора данных «Movieratings» представлен на рисунке 1.

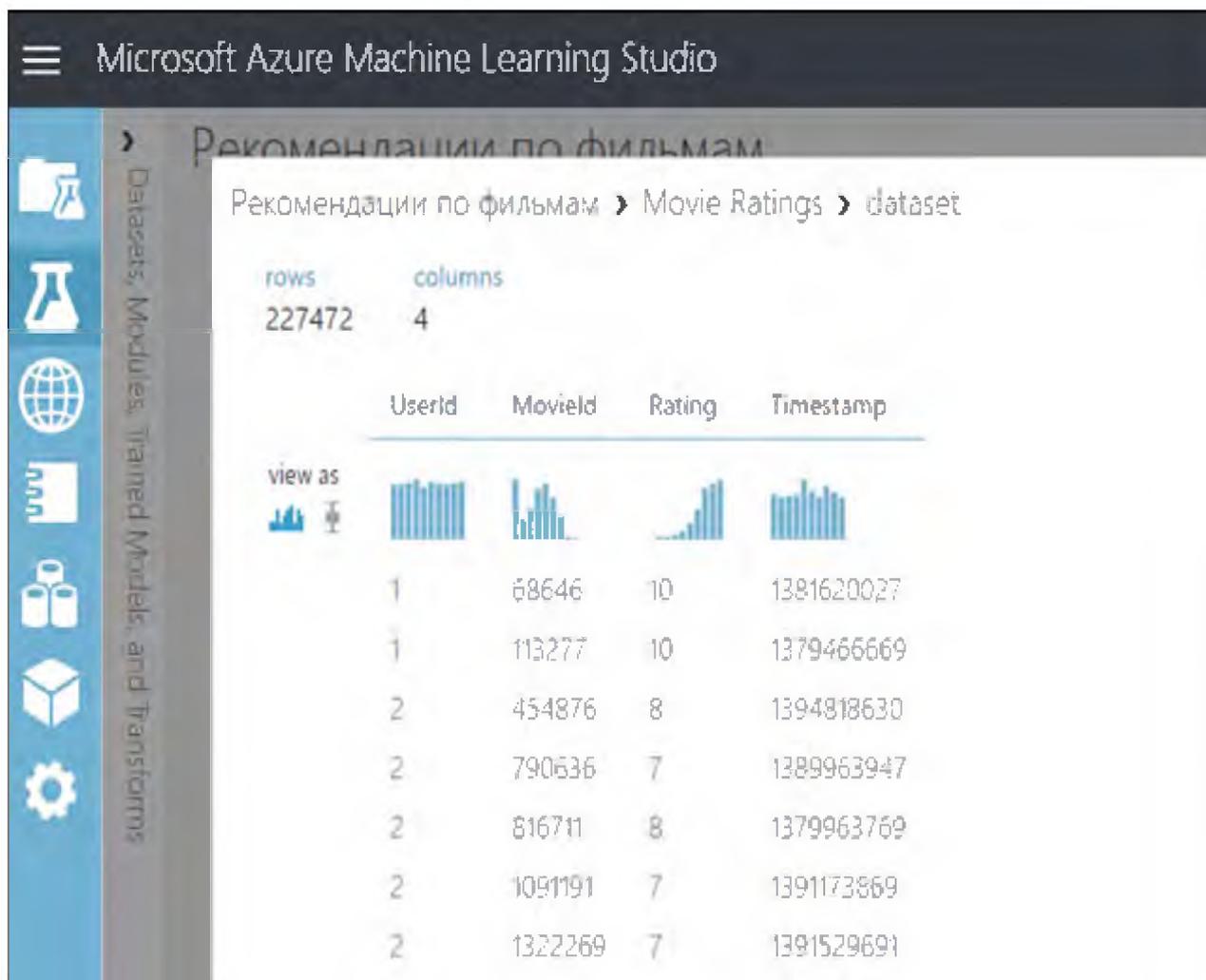


Рис. 1. Фрагмент содержимого набора данных «MovieRatings»

Второй набор данных «IMDB MovieTitles» - вспомогательный набор, который содержит названия фильмов для соответствующих уникальных идентификаторов. Фрагмент содержимого данного набора представлен на рисунке 2.

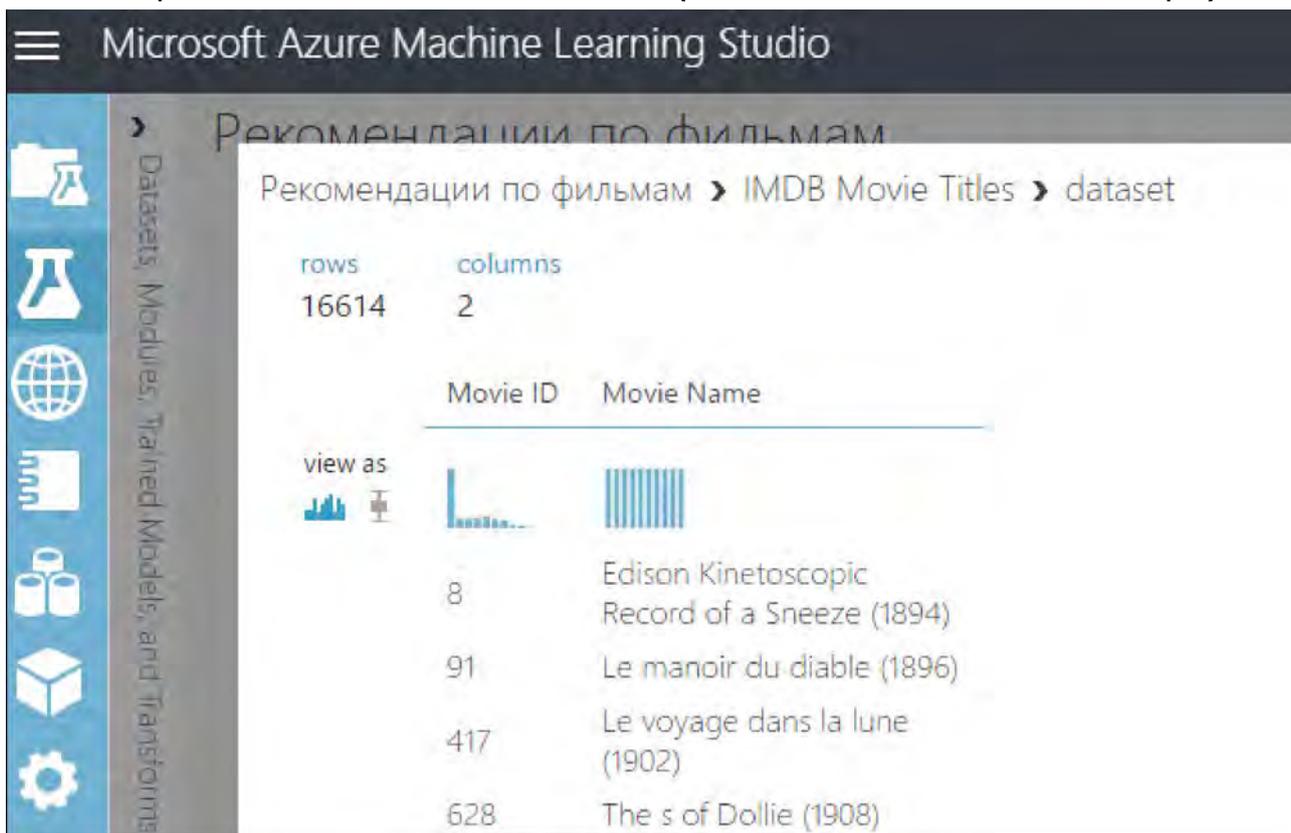


Рис. 2. Фрагмент содержимого набора данных «IMDB MovieTitles»

Далее с помощью модуля TrainMatchboxRecommender (Обучение системы рекомендаций Matchbox) машинного обучения Azure можно обучить построенную модель на основе Matchbox алгоритма. В результате мы получаем обученную рекомендационную модель.

Главная цель рекомендательных систем - предложить пользователю объект или несколько объектов, которые будут ему интересны. Объектами могут выступать любые предметы: книги, фильмы, рестораны, товары в магазинах и др. Под пользователями могут пониматься не только люди, или группы людей, а также любая другая сущность, которая обладает определёнными предпочтениями. Существуют два основных подхода, применяемых в рекомендательных системах: подход, основанный на содержимом, и коллаборативный подход. В первом подходе используются свойства объектов и пользователей, например, пользователи могут описываться такими свойствами как возраст, пол, страна. Объекты - свойствами автор, бренд, дата создания, страна производства. Во втором подходе используется неявно представленная информация, которая получается из матрицы оценок. MatchboxRecommender работает на основе синтеза этих подходов: коллаборативной фильтрации и фильтрации на основе содержимого. Поэтому этот алгоритм относится к гибридным рекомендательным подходам. Если пользователь является сравнительно новым в системе, качество предсказаний улучшается за счёт использования информации о самом пользователе и, таким образом, решается проблема «холодного старта». Однако, как только в системе накапливается достаточное количество информации об оценках данного пользователя, можно осуществлять персонализированные предсказания не только основываясь на данных о пользователе, но также на основе его оценок в системе [3].

Применив в эксперименте модуль TrainMatchboxRecommender (Обучение системы рекомендаций Matchbox), а затем Scorematchboxrecommender (Оценка системы рекомендаций

СЕКЦИЯ 5.1 ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА (СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ)

Matchbox) на выходе получим набор из 5 фильмов, которые предназначены для рекомендации пользователям (рис. 3).

User	Item 1	Item 2	Item 3	Item 4	Item 5
14887	Inception (2010)	One Flew Over the Cuckoos Nest (1975)	Star Wars Episode V - The Empire Strikes Back (1980)	The Raid 2 Berandal (2014)	There Will Be Blood (2007)
19075	Star Trek Into Darkness (2013)	No se Aceptan Devoluciones (2013)	Man of Steel (2013)	Elysium (2013)	
24752	Argo (2012)	My Week with Marilyn (2011)	The Necessary Death of Charlie Countryman (2013)	The Last Waltz (1978)	Our Idiot Brother (2011)
21707	Django Unchained (2012)	The Wolf of Wall Street (2013)	Capote (2005)	Behind the Candelabra (2013)	

Рис. 3. Полученные рекомендации

Для оценки точности построенной модели используются метрики качества ранжирования. Ранжирование - задача сортировки набора элементов из соображения их релевантности. В большинстве случаев релевантность понимается как отношение к некому объекту. Например, в задаче рекомендаций объект - это пользователь, элементы - тот или иной рекомендуемый материал, релевантность - вероятность того, что пользователь воспользуется данным материалом. В рассмотренном эксперименте была использована метрика NDCG (NormalizedDiscountedCumulativeGain). Значение данной метрики получено как 0.954984, что является достаточно хорошим результатом.

На рис.4 представлен построенный в AzureMachineLearning эксперимент.



Рис. 4. Построенный эксперимент в AzureMachineLearning

Выводы

В работе рассмотрены возможности AzureMachineLearningstudio. Построен эксперимент, в котором получена обученная модель для выдачи рекомендаций по фильмам для заданных пользователей. Посредством инструментов AzureMachineLearning данная модель может быть опубликована как веб-сервис, который будет использоваться в разрабатываемой рекомендательной системе.

Библиографический список

1. Alpaydin E. Introduction to Machine Learning, Third Edition. Изд-во: «The MIT Press», 2014.
2. Machine Learning Studio: Algorithm and Module Help [Электронный ресурс] – Режим доступа: <https://msdn.microsoft.com/en-us/library/azure/dn905974.aspx> – Дата доступа: 27.01.2017.
3. Stern D.H., Herbrich, R., Graepel T. Matchbox: Large Scale Bayesian Recommendations.- Proceedings of the 18th International World Wide Web Conference, 2009. 111-120 с.

Y.A. Skudnyakov, V.A. Moroz, N.N. Hurski

**USE OF MACHINE LEARNING TOOL ‘AZURE MACHINE LEARNING’
IN RECOMMENDER SYSTEMS**

Belarusian state University of Informatics and Radioelectronics, Minsk
Belarusian national technical University, Minsk

This article presents the main opportunities of Azure Machine Learning environment to create a model for use in the developed recommender system. Test experiment was build, which received training model for getting films recommendations to specified users.

Key words: recommender systems; machine learning; Azure Machine Learning; ranging; relevance.