

таком подходе система основывается чаще всего на пользователях. Алгоритмы рекомендаций позволяют посоветовать книги, используя ключевые слова или фразы в ней. Можно сделать вывод, что для разрабатываемого сервиса рекомендаций художественной литературы оптимальным вариантом является метод анализа содержимого объекта. Ключевые слова есть в каждой книге, а пользователи не всегда ставят оценки. При использовании коллаборативной фильтрации необходимо учитывать, что некоторые пользователи намеренно занижают или завышают оценку объектам.

ЛИТЕРАТУРА

1. Гомзин А.Г., Коршунов А.В. Системы рекомендаций: обзор современных подходов. – Труды Института системного программирования РАН, 2012 т.22. – С. 401-417.
2. Воронцов К.В. Методы коллаборативной фильтрации и тематического моделирования. URL: <http://www.machinelearning.ru/wiki/images/9/95/Voron-ML-CF.pdf> (дата обращения: 28.03.2016).
3. Имхонет – рекомендательный сервис. URL: <http://imhonet.ru/about/> (дата обращения: 28.03.2016).

ОПРЕДЕЛЕНИЕ АССОЦИАТИВНЫХ ПРАВИЛ МЕЖДУ КОЭФФИЦИЕНТАМИ РАСЧЕТНОЙ СКОРОСТИ

И.А. Петрухина

(г. Томск, Томский политехнический университет)

e-mail: irk-2009@yandex.ru

DETERMINATION OF ASSOCIATION RULES BETWEEN CRITERIA OF DESIGN SPEED

I.A. Petrukina

(Tomsk, Tomsk Polytechnic University)

The following article represents method of Data Mining such as market basket analysis. This method is used for finding association rules between criteria of design speed on a highway. As the result were obtained useful, trivial and incomprehensible rules.

Key words: data mining, market basket analysis, association rules, design speed.

Ассоциативные правила. Одной из наиболее распространенных задач Data Mining является исследование взаимосвязей между различными событиями для того, чтобы обнаружить ассоциации, зависимости между ними. Такой анализ называют анализом рыночной корзины, а полученные результаты – ассоциативными правилами [1]. Их основным достоинством является доступное восприятие человеком в форме логических конструкций «если..., то...», но не все правила представляют интерес. Выделяют три вида правил [2]:

- полезные – содержат действительную, ранее неизвестную информацию;
- тривиальные – содержат действительную, уже известную информацию;
- непонятные – содержат информацию, которая не может быть объяснена.

Структура данных. Целью данной работы является определение ассоциативных правил между коэффициентами обеспеченности расчетной скорости на автомобильной дороге. Входными данными являются участки автомобильной дороги, для каждого из которых определены 10 частных коэффициентов расчетной скорости K_{pcj} и нормативный показатель транспортно-эксплуатационного состояния $KП_n$. Данные K_{pcj} описывают различные характеристики автомобильной дороги [3]:

- K_{pc1} учитывает влияние на транспортно-эксплуатационное состояние дороги ширины проезжей части, количества полос и интенсивности движения.
- K_{pc2} учитывает влияние ширины и типа укрепления обочин.
- K_{pc3} зависит от интенсивности и состава транспортного потока на дороге.
- K_{pc4} определяют по величине продольного уклона и фактического расстояния видимости поверхности дороги при движении на подъеме.
- K_{pc5} определяют по величине радиуса кривой в плане и уклона виража.
- K_{pc6} учитывает ровность покрытия в продольном направлении.
- K_{pc7} учитывает влияние коэффициента сцепления покрытий автомобильных дорог.
- K_{pc8} показывает состояние покрытия дорожного полотна и прочности дорожной одежды.
- K_{pc9} рассчитывается в зависимости от величины колеи покрытий дорог.
- K_{pc10} определяют на основе сведений о ДТП и величины коэффициента относительной аварийности.

Если все $K_{pcj} \geq KП_n$, то состояние дороги считается нормативным, иначе же назначаются ремонтные работы. Как правило на одном участке присутствует несколько неудовлетворительных характеристик, в связи с чем может быть назначено несколько различных видов работ. Выявление наборов коэффициентов, связанных друг с другом, позволит назначать более эффективные ремонтные работы на данной автомобильной дороге.

Поиск правил. Определение ассоциативных правил осуществлялось с помощью алгоритма Apriori в аналитической платформе Deductor. Данные были нормализованы, а в качестве идентификатора выступал километраж дороги. В результате были получены следующие правила, представленные в таблице 1. Тривиальными считались те правила, которые описаны в [3].

Таблица 1 – Ассоциативные правила.

Полезные	Тривиальные	Непонятные
если K_{pc1} , то K_{pc4}	если K_{pc8} , то (K_{pc7}, K_{pc10})	если K_{pc5} , то $(K_{pc1}, K_{pc3}, K_{pc4})$
если K_{pc5} , то K_{pc3}	если K_{pc3} , то K_{pc4}	
если K_{pc3} , то (K_{pc5}, K_{pc1})	если K_{pc3} , то K_{pc1}	
	если K_{pc3} , то K_{pc5}	

Анализ полученных правил позволяет прийти к выводу, что многие из них тривиальны. Например, правило «если K_{pc8} , то (K_{pc7}, K_{pc10}) » говорит нам о том, что неудовлетворительное состояние покрытия дорожного полотна приводит к ухудшению сцепных качеств и к большей аварийности на данном участке дороги. Все полученные полезные правила не так очевидны, но понятны. А вот правило «если K_{pc5} , то $(K_{pc1}, K_{pc3}, K_{pc4})$ » (рис. 1), утверждающее, что недостаточный радиус кривой в плане влияет на видимость и интенсивность движения, можно назвать непонятным, так как его сложно объяснить. Для лучшего понимания требуется дополнительный анализ.

№	Номер правила	Условие	Следствие	Поддержка		Достоверность	Лифт
				Кол-во	%		
1	2	KPC5	KPC1	1633	45,11	47,53	1,029
			KPC3				
			KPC4				
2	1	KPC5	KPC1	1633	45,11	47,53	1,029
			KPC4				

Рисунок 1 - Непонятные правила.

Выводы. В данной работе были произведен поиск ассоциативных правил на наборе данных, характеризующих транспортно-эксплуатационное состояние автомобильной дороги. Анализ покупательской корзины, как один из инструментов Data Mining, позволяет эффективно обрабатывать большие объемы данных и находить необходимую информацию. Обнаруженные связи между различными частными коэффициентами расчетной скорости K_{pcj} позволяют выдвинуть предположения для построения дерева решений назначения видов ремонтных работ на данном участке дороги. Сравнение полученных правил с установленными в отраслевых дорожных нормах позволило судить об их тривиальности и полезности. Получение непонятных правил не является бесполезным, так в них могут скрываться глубокие знания, но их использование без дальнейшего изучения недопустимо.

Таким образом, использование ассоциативных правил позволяет найти на этапе предварительного анализа возможные скрытые зависимости и связи. Также ассоциативные правила позволяют определить наиболее часто встречающиеся наборы критериев и отбросить несущественные.

ЛИТЕРАТУРА

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика – от данных к знаниям. - СПб.: Питер, 2013. – 704 с.
2. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – СПб.: БХВ-Петербург, 2007. – 384 с.
3. ОДН 218.0.006-2002 Правила диагностики и оценки состояния автомобильных дорог. / Росавтодор Минтранса России, - М.: Информавтодор, 2002.-140 с.

ОБРАБОТКА РЕСУРСОВ УЗЛОВ В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ

Пилецкий А.А., Нанзатов А.Ж., Ботыгин И.А.
(г. Томск, Томский политехнический университет)
e-mail: aap47@tpu.ru

PROCESSING UNIT RESOURCES IN A DISTRIBUTED COMPUTING SYSTEMS

Piletskiy A.A., Nanzatov A.Z., Botygin I.A.
(Tomsk, Tomsk Polytechnic University)

The described program is designed for collecting, storing and processing data of distributed file system.

Keywords: distributed file system, data warehouse, grid computing, agent.