

## ТЕХНОЛОГИЧЕСКИЕ ПРОЦЕССЫ ТРАНСПОРТНОГО КОМПЛЕКСА

УДК 656.224:629.44

*А. А. ЕРОФЕЕВ, кандидат технических наук, Белорусский государственный университет транспорта, г. Гомель*

### ИСПОЛЬЗОВАНИЕ АЛГОРИТМА CLOPE В СИСТЕМЕ РАЗРАБОТКИ НОРМАТИВНОГО ГРАФИКА ДВИЖЕНИЯ Поездов

Проанализирован нормативный график движения поездов (НГДП). Установлено, что в течение года НГДП может не соответствовать фактическим размерам движения поездов. Рассмотрены технологии повышения соответствия между НГДП и фактическими размерами движения поездов. С целью повышения соответствия ГДП заявкам перевозчиков к параметрам системы поездной работы предложено формирование многослойного адаптивного ГДП (МАГ). Определены задачи разработки МАГ, решение которых требует применения технологий искусственного интеллекта. Предложено рассчитывать оптимальное количество слоев МАГ с использованием модифицированной версии алгоритма CLOPE. В результате моделирования установлено, что на количество слоев МАГ оказывают влияние уровень загрузки участка, разнообразие выделяемых категорий поездов, доля использованных ниток за прошлый год и скорость доставки.

В соответствии с [1] ГДП – нормативно-технический документ оператора инфраструктуры, устанавливающий организацию движения поездов всех категорий на участках инфраструктуры железнодорожного транспорта, графически отображающий следование поездов на масштабной сетке в условные сутки, подразделяемый на нормативный (на плановый год), вариантный (в отдельные периоды времени) и оперативный (на текущие плановые сутки) график движения поездов.

Нормативный ГДП разрабатывается один раз в год на прогнозные среднесуточные размеры движения месяца максимальной нагрузки. Однако аналитические исследования ГДП на участках Белорусской железной дороги (рисунок 1) позволяют сделать вывод, что имеются существенные отклонения исполненных размеров движения от нормативных.

Например, для анализируемого однопутного участка А при нормативных размерах движения, равных 15 поездов в сутки, среднемесячные исполненные размеры составляют от 7,3 до 10 поездов. При этом в разные месяцы размеры движения нечетных и четных поездов могут отличаться друг от друга как в большую, так и в меньшую сторону. Анализ фактических размеров движения за сутки одного месяца позволил выявить, что при нормативных размерах движения 15 пар поездов в сутки исполненные размеры движения изменялись от 5,12 до 16,74 поездов в нечетном направлении и от 5,37 до 13 поездов – в четном, т. е. даже в течение одного месяца разница между минимальными и максимальными размерами движения может превышать 320 %.

На практике существует два основных способа повышения соответствия между нормативными и фактическими размерами движения поездов.

1 Нормативный ГДП является основой, которая используется поездным диспетчером при разработке графиков движения на плановые периоды (как правило, на 3–6 часов) (традиционно использовался на железных дорогах СССР). Данный способ позволяет эффективно учитывать возникающие отклонения структуры и мощности поездопотока от заложенных в нормативном ГДП. При этом за основу принимается гипотеза о незначительных отклонениях фактических вагонопотоков от плановых значений в больших периодах времени (характерно для устойчивых параметров вагонопотоков СССР).

Однако в современных условиях разработка диспетчерских вариантов ГДП имеет существенный недостаток. Технология разработки ГДП предполагает, что график на основании ПФП увязывается на всех участках следования поездного назначения. ГДП также должен быть увязан с технологией работы сортировочных, участковых, грузовых и пассажирских станций, системой эксплуатации локомотивов. В силу ограниченности полигона управления ДНЦ не в состоянии обеспечить решение данной задачи в условиях значительных отличий логистических характеристик прогнозных вагонопотоков от нормативных значений. По этой причине при эффективном пропуске поездов по участку возникают риски простоя поездов на технических станциях (отсутствие взаимовязанной нитки ГДП на последующем участке, отсутствие локомотива для отправления

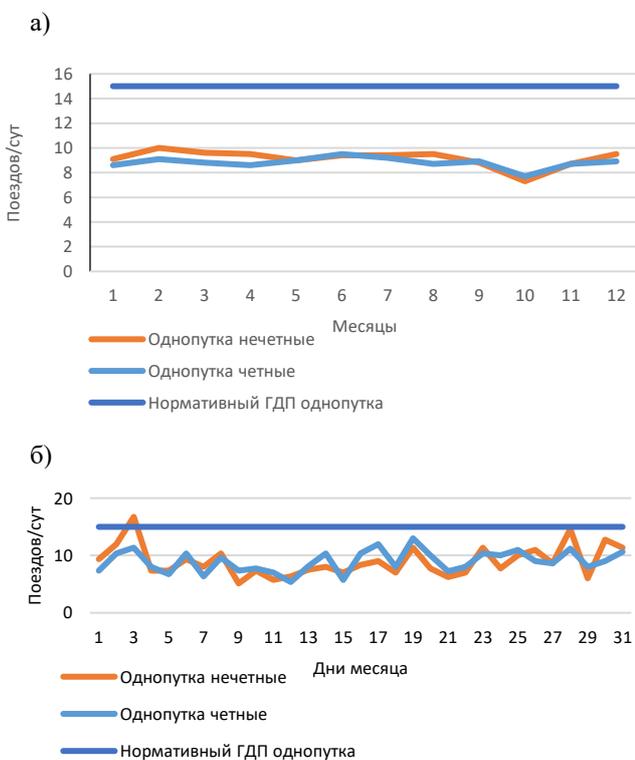


Рисунок 1 – Изменение интенсивности движения грузовых поездов на однопутном участке потока:  
а – в течение года; б – в течение месяца

поезда на следующий участок). Как следствие, маршрутная скорость продвижения вагонопотока снижается, возникают дополнительные логистические издержки у операторов инфраструктуры (увеличение потребностей в станционных путях), перевозчиков (снижение производительности работы локомотивов и локомотивных бригад) и клиентов (замедление доставки грузов).

2 Увеличение частоты разработки нормативного ГДП. Данный способ предполагает разработку НГДП несколько раз в год (как правило, от 2 до 4) (используется в настоящее время на Белорусской железной дороге и ряде зарубежных железных дорог). При таком подходе появляется возможность учесть годовые неравномерности в структуре и мощности поездопотоков, увязать ГДП с ПФП, установить размеры передачи поездов по международным стыковым пунктам и специализацию грузовых поездов. Однако проблема компенсации внутримесячной неравномерности остается нерешенной и в дальнейшем переносится на уровень ДНЦ со всеми присущими способу 1 недостатками.

Таким образом, необходимо формирование адаптивной системы разработки НГДП, которое позволит учитывать требования клиентов и перевозчиков к параметрам системы доставки грузов и будет способствовать снижению общетранспортных издержек в логистических цепях.

В соответствии с действующей технологией поезда на графике прокладываются последовательно по их категориям. Вначале прокладываются пассажирские поезда в соответствии с предварительно выбранной схемой их обращения на данном направлении, затем региональные, ускоренные грузовые и далее остальные грузовые поезда, т. е. разработка ГДП предусматривает этапность и приоритеты. График движения транзитных грузовых поездов, следующих без переработки на попутных технических станциях, составляется сквозным на всём пути их следования. Это обеспечивает согласованную работу смежных участков, повышает маршрутную скорость движения поездов и позволяет лучше увязать оборот поездных локомотивов. Следовательно, ГДП должен разрабатываться не отдельно по участкам, а в целом для всего направления. В связи с этим совокупность количественной и временной неопределенности приводит к *процессной неопределенности*.

В итоге в системе пропуска поездов возникают «конфликты» при построении ниток поездов различных категорий и приоритетов. Они возникают как на этапе прокладки ниток ГДП, так и на этапе согласования ГДП между участками, а тем более дорогами. В отдельных случаях такой многоуровневый конфликт не позволяет использовать стандартные последовательности и алгоритмы разработки.

Одним из основных параметров, характеризующих качество разработки ГДП, является участковая скорость. На рисунке 2 приведен анализ распределения технических и участковых скоростей движения отдельных грузовых поездов. Как видно из рисунка 2, а, на исследуемом однопутном участке А техническая скорость движения поездов изменяется в пределах от 35 км/ч до 63 км/ч, принимая наиболее частое значение 51 км/ч для 27 % поездов. При этом участковая скорость движения поездов изменяется в пределах от 35 до 59 км/ч, принимая наиболее частые значения 36 км/ч для 13 % поездов и 47 км/ч в 14 % случаев.

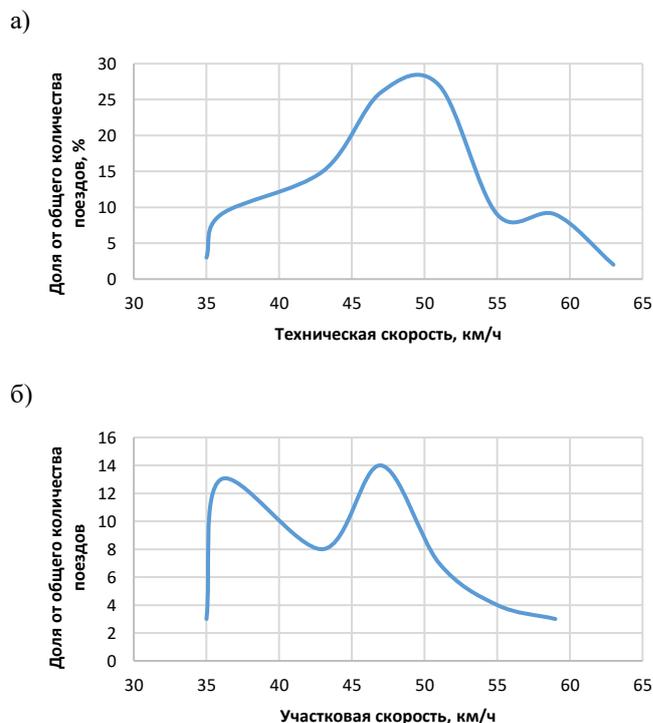


Рисунок 2 – Распределение технической (а) и участковой (б) скоростей движения грузовых поездов на однопутном участке (нормативный ГДП)

Можно сделать вывод, что на участковую скорость движения грузовых поездов оказывают непосредственное влияние не только параметры поездопотока, но и условия прокладки нитки поезда в ГДП (количество остановок для обгонов и скрещений поездов).

Интеллектуальная разработка ГДП предполагает не только создание информационно-расчетных процедур и алгоритмов, но и изменение технологических подходов к разработке графика.

Одним из таких подходов является переход к системе формирования **многослойного адаптивного ГДП** (МАГ). Слой ГДП – это совокупность взаимоувязанных ниток ГДП, разработанных для сходной эксплуатационной обстановки и обеспечивающих пропуск установленного количества поездов.

Упорядоченная совокупность слоев образует ГДП, действующий в установленный период времени (год, месяц, сутки и др.). В зависимости от требуемых размеров движения поездов нормативный ГДП может актуализироваться путем комбинирования слоев или ниток различных слоев между собой. При этом важным условием разработки многослойного ГДП является отсутствие «конфликтов» между нитками различных слоев, а также заблаговременное установление приоритетов для ниток, принадлежащих различным слоям, т. е. разработка ГДП в ИСУПП предполагает процедуры не только непосредственной разработки, но и последующей актуализации ГДП в зависимости от складывающейся эксплуатационной обстановки (например, в ГДП заложена нитка пассажирского поезда, который назначается приказом Начальника железной дороги. В период отсутствия такого поезда в это время может назначаться грузовой, причем с перегонными временами, отличными от пассажирского).

В системе формирования МАГ используются ключевые принципы, заложенные технологией разработки

ГДП с выделением «твердых» расписаний грузовых поездов предусматривает [2]:

- учет сезонных или помесечных колебаний поездопотоков путем отнесения поездов к различным слоям, что позволяет в одном графике совмещать несколько вариантов размеров движения поездов;

- групповую специализацию расписаний, при которой по одной нитке графика в разные сутки могут следовать поезда разных назначений (в частности, транзитные или в расформирование, сквозные поезда или отправительские маршруты);

- размещение на разных слоях ниток поездов, имеющих различный приоритет, но проложенных в одинаковые временные интервалы (конфликтующие нитки с различными перегонными временами хода).

МАГ предполагает организацию движения грузовых поездов как по постоянному, так и по свободному расписанию. Однако ключевым отличием от технологии «твердых» расписаний является динамически изменяемое количество слоев и правила отнесения поездных заявок к соответствующему слою.

При разработке МАГ для каждой нитки указывается периодичность ее использования и признак принадлежности соответствующему слою. При формировании актуального ГДП (на конкретные сутки) могут использоваться как все нитки одного слоя, так и часть из них. При этом в первую очередь должны использоваться нитки более «высокого» слоя, а затем, при их недостатке, дополняться нитками «нижнего» слоя.

Таким образом, при разной эксплуатационной обстановке один и тот же график может использоваться без пересоставления, а наиболее часто используемые расписания обеспечивают более высокую маршрутную (участковую) скорость и меньшее число остановок для обгонов и скрещений поездов. При этом соблюдается преемственность между отдельными вариантами за счет того, что поезда, которые включены в ГДП, составленный на наименьшие размеры движения, внесены также и в остальные варианты.

Задачами интеллектуализации ИСУПП следует считать:

- 1) идентификацию логистических характеристик поездопотоков и определение параметров соответствующих ниток ГДП;
- 2) определение количества слоев ГДП;
- 3) определение количества поездов, относящихся к каждому из слоев ГДП;
- 4) разработку многослойного ГДП [3];
- 5) актуализацию ГДП [4].

В данной статье рассмотрено решение второй и третьей задач. Эти эксплуатационные задачи можно отнести к задачам обработки массивов данных Data Mining и последующей их кластеризации.

При такой постановке совокупность поездных заявок и их уникальные эксплуатационные характеристики можно отнести к категориальным данным (т. е. качественным характеристикам объектов, измеренным в шкале наименований). При использовании шкалы наименований решается задача, одинаковы или нет объекты относительно измеряемого признака.

В настоящее время существует множество методов для работы с категориальными данными, например, семейство иерархических кластерных алгоритмов, в которых вместо локальных критериев (попарное сравнение)

используются глобальные критерии. Для кластеризации поездных заявок предлагается использовать алгоритм CLOPE и его модифицированную версию [5], которые обеспечивают лучшее качество кластеризации в сравнении со многими иерархическими алгоритмами.

Формализуем задачу кластеризации поездных заявок для категориальных данных.

Задача кластеризации поездных заявок состоит в получении такого разбиения всего их множества, чтобы похожие оказались в одном кластере, а отличающиеся друг от друга – в разных кластерах. В качестве кластера будем рассматривать слой МАГ.

Пусть имеется множество поездных заявок (база транзакций  $D$ ), состоящая из множества транзакций – поездных заявок  $\{t_1, t_2, \dots, t_n\}$ . Каждая транзакция предполагает наличие определенных требований к условиям предоставления доступа (национальный перевозчик или нет, категория поезда, потребная скорость перевозки, заявленная частота отправления и др.) и может быть описана совокупностью свойств  $\{i_1, \dots, i_m\}$ . Множество кластеров  $\{C_1, \dots, C_k\}$  есть разбиение множества  $\{t_1, \dots, t_n\}$ , такое, что

$$C_1 \dots C_k = \{t_1, \dots, t_n\} \text{ и} \quad (1)$$

$$C_i \neq \emptyset \wedge C_i \cap C_j = \emptyset, \text{ для } 1 \leq i, j \leq k. \quad (2)$$

Каждый элемент  $C_i$  будет являться кластером (слоем графика);  $n, m, k$  – соответственно количество транзакций, количество объектов в базе транзакций и число кластеров.

Каждый кластер  $C$  имеет следующие характеристики:

$D(C)$  – множество уникальных поездных заявок;

$Occ(i, C)$  – частота включения поездной заявки  $i$  в слой  $C$ ;

$$S(C) = \sum_{i \in D(C)} Occ(i, C) = \sum_{i \in C} |t_i|;$$

$$W(C) = |D(C)|;$$

$$H(C) = S(C)/W(C).$$

Глобальный критерий качества распределения поездных заявок по слоям может быть выражен в виде функции стоимости  $Profit(C)$ :

$$Profit(C) = \sum_{i=1}^k |C_i| \sum_{i=1}^k G(C_i) \times |C_i| = \sum_{i=1}^k |C_i| \sum_i i = 1k W(C_i) r S(C_i) \times |C_i|, \quad (3)$$

где  $|C_i|$  – количество поездных заявок в  $i$ -м слое;  $k$  – количество слоев;  $r$  – положительное вещественное число, большее 1.

Параметр  $r$  является «коэффициентом отталкивания» (repulsion), регулирует уровень сходства поездных заявок внутри слоя и, как следствие, количество слоев МАГ. Данный коэффициент может корректироваться. Чем больше  $r$ , тем ниже уровень сходства и тем больше слоев будет сгенерировано.

Постановка задачи кластеризации МАГ алгоритмом CLOPE формулируется следующим образом: для заданных  $D$  и  $r$  найти разбиение  $C$ :  $Profit(C, r) \rightarrow \max$ .

С целью последующей обработки массив поездных заявок представляется в виде таблицы транзакций категориальных данных. Для этого для всей совокупности свойств поездных заявок формируются шкалы наименований.

Массив поездных заявок приводится к нормализованному виду. Он может иметь вид бинарной матрицы образов, как в ассоциативных правилах, так и представлять собой взаимно однозначное отображение между множеством уникальных объектов  $\{u_1, \dots, u_q\}$  таблицы и множеством целых чисел  $\{0, 1, 2, \dots, q-1\}$ .

Классический подход к задаче кластеризации предполагает полностью автоматический процесс без возможности задания каких-либо априорных данных/правил при решении задачи. Это приводит к тому, что повлиять на результат кластеризации никак нельзя, а итоговый результат во многом зависит от выбранного алгоритма кластеризации.

В реальных условиях при решении задачи разработки МАГ могут фигурировать различные критерии выбора приоритетного отнесения тех или иных поездных заявок к более «скоростным» слоям (обеспечение максимальной участковой скорости, приоритетный пропуск пассажирских поездов, ускоренный пропуск поездов национального перевозчика, обеспечение безостановочного пропуска поездов повышенной длины и др.).

В таких случаях возникает необходимость использования при кластеризации небольшого количества априорных данных, устанавливающих приоритеты для конкретной эксплуатационной обстановки. Такой подход получил название кластеризации с неполным обучением [6]. В рамках кластеризации с неполным обучением имеется некоторое априорное знание об анализируемых данных, например, априорные сведения о значимости некоторых параметров поездных заявок при их отнесении к определенному слою МАГ. Значимость атрибута выражается весовым коэффициентом атрибута. Каждому атрибуту  $a_i$  ставится в соответствие действительное положительное число  $w_i$  – весовой коэффициент. Чем больше значение весового коэффициента, тем больше вклад атрибута в формирование кластеров. В стандартном алгоритме CLOPE весовые коэффициенты всех атрибутов равны единице. Путем изменения исходные значения весовых коэффициентов можно формировать состав кластеров в зависимости от поставленных приоритетов [7].

В отличие от стандартного CLOPE в модифицированном алгоритме необходимо вычислять дополнительные характеристики кластера:

1)  $a_i$  – номер атрибута, которому принадлежит значение  $i$ ;

2)  $A(C)$  – множество уникальных значений атрибутов в кластере  $C$ .

Кроме того, предлагается изменить способ вычисления следующих характеристик кластера:

$$S(C) = |C| \sum_{i=1}^m w_i - \text{площадь кластера}; \quad (4)$$

$$W(C) = \sum_{i \in A(C)} w_{a(i)} - \text{ширина кластера}. \quad (5)$$

Рассмотрим пример кластеризации поездных заявок по слоям МАГ. Форма исходных данных представлена в таблице 1.

Таблица 1 – Исходные данные о поездных заявках

Поездная заявка	Категория поезда	Перевозчик	Скорость доставки	Вид перевозимого груза	Длина	Частота отправления	Доля использованных ниток за прошлый год	Сезонность	Постоянное или свободное расписание
1									
2									
...									
$i$									
...									
$D$									

В данной работе для оценки качества кластеризации предлагается использовать индекс оптимального соответствия. Вычисление данного индекса заключается в поиске такого инъективного отображения множества эталонных кластеров в множество кластеров итогового разбиения, при котором сумма пересечений эталонных кластеров и итоговых кластеров является максимальной. Рассмотрим процедуру вычисления индекса оптимального соответствия.

Пусть,  $A_1, \dots, A_k$  – эталонные кластеры, полученные в результате генерации исходного массива данных;  $k$  – эталонное количество кластеров;  $n_i$  – количество объектов в  $i$ -м кластере;  $R_1, \dots, R_k$  – кластеры (слои МАГ), полученные в ходе выполнения алгоритма;  $k'$  – количество кластеров (слоев МАГ), которое было определено алгоритмом кластеризации.

Для оценки качества кластеризации необходимо получить однозначное соответствие эталонного кластера  $A_i$  результирующему кластеру  $R_p$ . Множество индексов  $\{p_1, \dots, p_k\}$  должно быть подобрано таким образом, чтобы максимизировать сумму  $c_i$  – количества верно определенных объектов в эталонном кластере  $A_i$ . Исходя из этого

$$c_i = |A_i \cap R_{p_i}|; \bigcap_i p_i = \emptyset; \quad (6)$$

$$U_i p_i = \arg \max \sum_i c_i. \quad (7)$$

Определив количество объектов  $c_i$ , вычисляем индекс оптимального соответствия по формуле

$$I_{oc} = \frac{\sum_{i=1}^k \frac{c_i}{n_i}}{k}. \quad (8)$$

В формуле (8) числитель представляет собой сумму степеней корректности определения кластеров. Определим область значений индекса качества кластеризации. Наименьшее значение индекса качества кластеризации возможно в том случае, если алгоритм определяет каждый объект в отдельный кластер. Тогда значение индекса

$$I_{min} = \frac{\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}}{k} = \frac{\sum \frac{1}{n}}{k}. \quad (9)$$

Наибольшее значение индекса качества кластеризации возможно в том случае, если алгоритм точно определил количество кластеров и корректно определил состав кластеров. Степень корректности определения каждого кластера равна единице. Значение индекса при этом также равно единице:

$$I_{max} = \frac{\frac{n_1}{n_1} + \frac{n_2}{n_2} + \dots + \frac{n_k}{n_k}}{k} = \frac{k}{k} = 1. \quad (10)$$

Таким образом, область значений индекса качества кластеризации определена на отрезке  $\left[ \frac{\sum_{i=1}^k \frac{1}{n_i}}{k}; 1 \right]$ ; 1, где  $k$  – априорное количество кластеров.

Особенностью данного метода оценки является то, что степень корректности определения кластеров имеет равный вклад в значение индекса вне зависимости от размера определяемых кластеров.

## Список литературы

**Результаты.** На основании тестовой выборки поездных заявок выполнен расчет оптимального количества слоев МАГ с использованием улучшенного алгоритма CLOPE. В качестве приоритетных параметров, для которых весовые коэффициенты  $w_i$  принимались равными 2, были выбраны: категория поезда, частота отправления по нитке, постоянное или свободное расписание. Для остальных параметров весовые коэффициенты были приняты равными 1.

В результате установлено, что для однопутных участков с уровнем загрузки до 50 % достаточно двух слоев МАГ; для однопутных с уровнем загрузки свыше 50 %, а также для двухпутных участков с загрузкой до 40 % – три слоя; для двухпутных участков с загрузкой свыше 40 % рекомендуется использовать четыре слоя.

Кроме того, в результате моделирования установлено, что на количество слоев кроме уровня загрузки участка непосредственное влияние оказывают разнообразие выделяемых категорий поездов, доля использованных ниток за прошлый год и скорость доставки.

Использование улучшенного алгоритма CLOPE при разработке МАГ позволяет существенно упростить процедуру классификации поездных заявок и обеспечить реализацию приоритетных целей оператора инфраструктуры при разработке графика.

Получено 20.05.2020

**А. А. Erofeev.** Using the CLOPE algorithm in the development system standard train schedule.

As a result of the analysis, it was established that during the year the normative train schedule (NGDP) may not correspond to the actual size of train traffic. The technologies for improving the correspondence between the NGDP and the actual size of train traffic are considered. In order to increase the compliance of the traffic flow with the carriers' requests to the parameters of the train operation system, it is proposed to form a multilayer adaptive traffic flow (MAG). The tasks of MAG development are determined, the solution of which requires the use of artificial intelligence technologies. It is proposed to calculate the optimal number of MAG layers using a modified version of the CLOPE algorithm. As a result of the simulation, it was found that the number of MAG layers is influenced by the level of the section load, the variety of selected categories of trains, the share of lines used for the last year and the speed of delivery.

1 Правила технической эксплуатации железной дороги в Республике Беларусь : утв. постановлением М-ва трансп. и коммуникаций Респ. Беларусь 25.11.2015 № 52. – Минск, 2016. – 568 с.

2 Инструкция по организации поездной работы при отпращивании грузовых поездов по твердым ниткам графика : утв. вице-президентом ОАО «РЖД» С. В. Козыревым 19 дек. 2006 г. – М., 2006. – 47 с.

3 Комплекс прикладных программ для разработки графика движения поездов / В. Г. Кузнецов [и др.] // Вестник ВНИИЖТа. – 2007. – № 4. – С. 25–31.

4 **Erofeev, A. A.** Multi-criteria evaluation of management decisions in the intellectual system of transportation management / A. A. Erofeev // Открытые семантические технологии проектирования интеллектуальных систем : сб. науч. трудов. – Вып. 3 ; редкол. В. В. Голенков [и др.]. – Минск : БГУИР, 2019. – С. 205–208.

5 **Yang, Y.** CLOPE: A fast and Effective Clustering Algorithm for Transactional Data In Proc of SIGKDD'02 / Y. Yang, H. Guan, J. You. – Edmonton, Alberta, Canada, 2002.

6 **Girra, N.** Unsupervised and Semi-supervised Clustering: a Brief Survey / N. Girra, M. Crucianu, N. Boujemaâ // Proc. of 7th ACM SIGMM international workshop on Multimedia information retrieval. – 2004.

7 **Бильгаева, Л. П.** Модификация алгоритма кластеризации CLOPE / Л. П. Бильгаева, З. Г. Самбялов // Фундаментальная наука и технологии – перспективные разработки : сб. ст. II Междунар. науч.-практ. конф. – Т. 2. – М. : НИЦ «Академический», 2013. – С. 86–90.