

**КЛАСТЕРИЗАЦІЯ МОРСКИХ СООБЩЕНИЙ***Вильский Г.Б.,**Николаевский политехнический институт*

*В статье рассмотрены проблемы кластерного анализа морских сообщений, приведена классификация методов кластеризации и на их основе разработаны кластеры потоков данных на водных путях информационно-безопасности движения судна.*

*Ключевые слова: кластерный анализ, потоки данных, информационная безопасность движения судна.*

**Постановка проблемы в общем виде.** Реалии судоходства обозначили роль и значение морских сообщений в безопасности судовождения. Изучение морских инцидентов за последние 10 лет указывает на ряд проблем по обеспечению оперативными данными об обстановке на маршрутах движения судов. В показателях случаев от получаемых с береговых систем управления движением судов (СУДС) преувалируют данные с недостаточностью сведений с угрожающими факторами. Действующие СУДС формируют информационные потоки фрагментарно, регламентно и подают их судоводителям бессистемно, а это, как свидетельствует практика, не позволяет отдавать верные команды управления судном. Такое положение приводит к неадекватной оценке параметров в морских сообщениях и снижает безопасность управления судном в реальных навигационных условиях [1]. Поставленная проблема управления судном обусловлена недостаточностью научных разработок в области информационного обеспечения судовождения. Актуальность их выполнения подтверждается п. 1 («Развитие технологий поддержки принятия стратегических решений»), п. 3 («Развитие технологий ситуативного управления во время принятия управленческих решений») и п. 7 («Развитие навигационных систем разного назначения») постановления КМУ от 12 марта 2012г. № 294 «Некоторые вопросы определения среднесрочных приоритетных направлений инновационной деятельности общегосударственного уровня на 2012-2016 годы».

Анализ последних достижений и публикаций. Изучение организации работ и функциональных взаимосвязей между предприятиями морехозяйственного комплекса Украины показывает их построение в соответствии с механизмом теории кластеров. Данное утверждение подтверждается действующей в виде разветвленно-структурированных цепочек вертикали морской администрации страны и системой циркуляции потоков сведений между береговыми и морскими подвижными объектами. В виде морского кластера навигационной безопасности представлена совокупность технических средств, каналов и сетей передачи данных для управления судами, входящих в состав интерактивных головных и региональных центров СУДС [2]. Отдельные специализированные

предприятия морской отрасли при формировании и передаче навигационных сообщений используют логистические методы построения информационного пространства водных путей с автоматической кластеризацией массивов параметров движения судов.

На механизме кластерной теории в судоходстве решен ряд задач безопасного управления судном. В работе [3] приведена новая спутниковая система навигационно-информационной технологии в судовождении. Её исследование с учётом дефиниции «кластер» и реализованное оплачивание позволило перейти к подготовке карт пространственных полей ошибок обсерваций в динамическом режиме судна и таким образом обеспечить судовождение технологией высокоточной навигации. Известные схемы обмена данными между судами и береговыми службами построены по кластерной технологии и реализуются за счёт совместимости аппаратуры спутниковых навигационных систем, телевизионных систем высокого разрешения и транспондеров АИС. В кластерной концепции информационного обеспечения проектов VTMISS и POSEIDON содержится интеграция морских сообщений от береговых и спутниковых систем связи [4]. Региональными СУДС и постами регулирования движения судов формируется информационное пространство и проводится вероятностная аналитика высокоточных параметров от обращенной радиолокационной системы. При этом наблюдение за движением судов обеспечивается кластерами непрерывных зон радиолокационного и транспондерного контроля [5, 6, 7]. Процедуры оценки угрозы ситуационного возмущения при опасном сближении судов предполагают использование классификации вариантов расхождения, в которой информационные взаимосвязи процесса маневрирования подчинены кластерному механизму. Благодаря кластерному представлению, путем ранжирования и показа иерархии морских сведений, авторы исследований [8, 9], предложили модель оптимизированной структуры системы управления судном. Несмотря на опосредованное представление кластерных принципов в морской отрасли, в отечественной и зарубежной морской литературе не представлена дефиниция морского кластера информационной безопасности движения судна. В связи с этим возникает необходимость использования кластерного механизма для систематизации выходных данных СУДС. Приведенный обзор научно-практических работ характеризует опосредованное использование в мореплавании кластерной теории. Важность кластеризации морских сообщений в информационном обеспечении СУДС определяет актуальность научного исследования.

**Цель статьи.** Показать важность для морской практики технологии кластерного описания морских сообщений. Разработать кластеры признаков потоков данных на водных путях и информационной безопасности движения судна.

**Изложение материала исследования.** В информационном пространстве водных путей морские сообщения делятся на пять смысловых

категорий: о бедствии и опасности, навигационные, гидрометеорологические, эксплуатационные и частная переписка. Поток данных с такими смысловыми категориями автоматически классифицируются (кластеризируются) по признакам. Целью кластерного анализа морских сообщений является образование схожих по заданным критериям групп объектов (кластеров) для анализа информационных потоков водных путей. Методами кластерного анализа решаются следующие задачи судовождения:

- проведение классификации средств навигационного оснащения (СНО) по признакам. Решение такой задачи, как правило, приводит к углублению знаний о совокупности классифицируемых объектов;
- проверка и поиск существующего локального кластера безопасности;
- построение автоматической системы классификации для маловероятностных событий и установление внутренних взаимосвязей в сведениях;
- оптимизация данных – путём сжатия исходной выборки при её избыточности с оставлением одного типичного представителя от каждого кластера.

Кластерный анализ выполняется поэтапно и включает: отбор выборки морских сообщений; определение признаков пространства; вычисление значений той или иной меры сходства между объектами; создание групп сходных объектов; проверка достоверности результатов. Показатели потоков с данными должны отвечать нормальному вероятностному распределению, быть устойчивыми к влиянию случайных факторов, удовлетворять однородности в выборке и не коррелировать между собой. При описании морских сообщений в виде объектов кластеризации приводится метрическое пространство, на котором каждому объекту присущи соответствующие наборы характеристик и определены расстояния между ними. Признаки могут быть числовыми или нечисловыми, а степень сходства объектов выборки в метрическом пространстве учитывается матрицей расстояний (различий) между объектами до 1. Существуют агломеративные, дивизимные и итеративные методы кластерного анализа (автоматической классификации) морских сообщений. При агломеративном методе [10] предполагается, что каждое морское сообщение в начале анализа является отдельным кластером и производится группировка схожих сообщений на основании матрицы мер сходства по кластерам. Вычисляются центры кластеров, каждому сообщению присваивается свой номер кластера для дальнейшего использования кластеров при автоматизированных расчетах без повторной кластеризации данных. Дивизимные методы расчленяют группы (кластеры) на отдельные объекты. Каждый метод реализуется при помощи собственного алгоритма. При итеративных методах кластеры формируются исходя из задаваемых условий градации, которые могут быть изменены операторами СУДС для достижения желаемого качества потоков данных управления судном. В

итеративных методах нашли применение метод – средних, метод поиска сгущений и метод иерархической кластеризации. Метод иерархической кластеризации приводит к построению дерева – дендрограммы, позволяющей визуально определять принадлежность объекта к данному типу кластеров. Преимущество итеративных методов определяется их быстродействием в обработке больших массивов исходных данных. При формировании кластеров информационных потоков водных путей возможны для использования: эвристические графовые алгоритмы; функциональные качества кластеризации; статистические алгоритмы и алгоритмы иерархической кластеризации.

Общая картина кластеров информационных потоков на водном пути по признакам приведена на рис. 1. Восемь признаков информационных потоков водных путей – направленность, происхождение, назначение, защищённость, коммуникативность, режим работы, периодичность и носитель информации составляют основу кластерных полей морских сообщений. Каждый из отмеченных признаков может образовывать свои разветвления с детальным конкретным наполнением, характеризующим процессы формирования, передачи, отображения, циркуляции потоков информации и их защищённость. Возможное дополнительное количество классификационных признаков в разветвлениях может колебаться, как правило, в пределах от двух до четырёх. Между разветвлениями могут образовываться межкластерные связи, дающие полное представление о характеристике информационного поля на водном пути. Взаимную зависимость возможных предпосылок и угрозных факторов характеризует специальный кластер информационной безопасности управления судном. Кластер показывает расстояния между гроздьями состояния обстановки на мостике судна, угрозных факторов, свершившихся аварий и автоматически упорядочивает их связь в сравнительно однородных классификационных группах. Кластер информационной безопасности движения судна, показанный на рис. 2, открывает возможность селективного сбора данных, способствует выполнению расчётов угроз информационной безопасности и моделированию вероятностных распределений опасностей на участках водного пути.

Исходя из принципиальных положений кластерного анализа, задача кластеризации информационных потоков водных путей заключается в градации имеющийся морских сообщений на ряд не пересекающихся подмножеств. Необходимо, чтобы каждый кластер состоял из объектов, близких по метрике, а объекты разных кластеров существенно отличались.

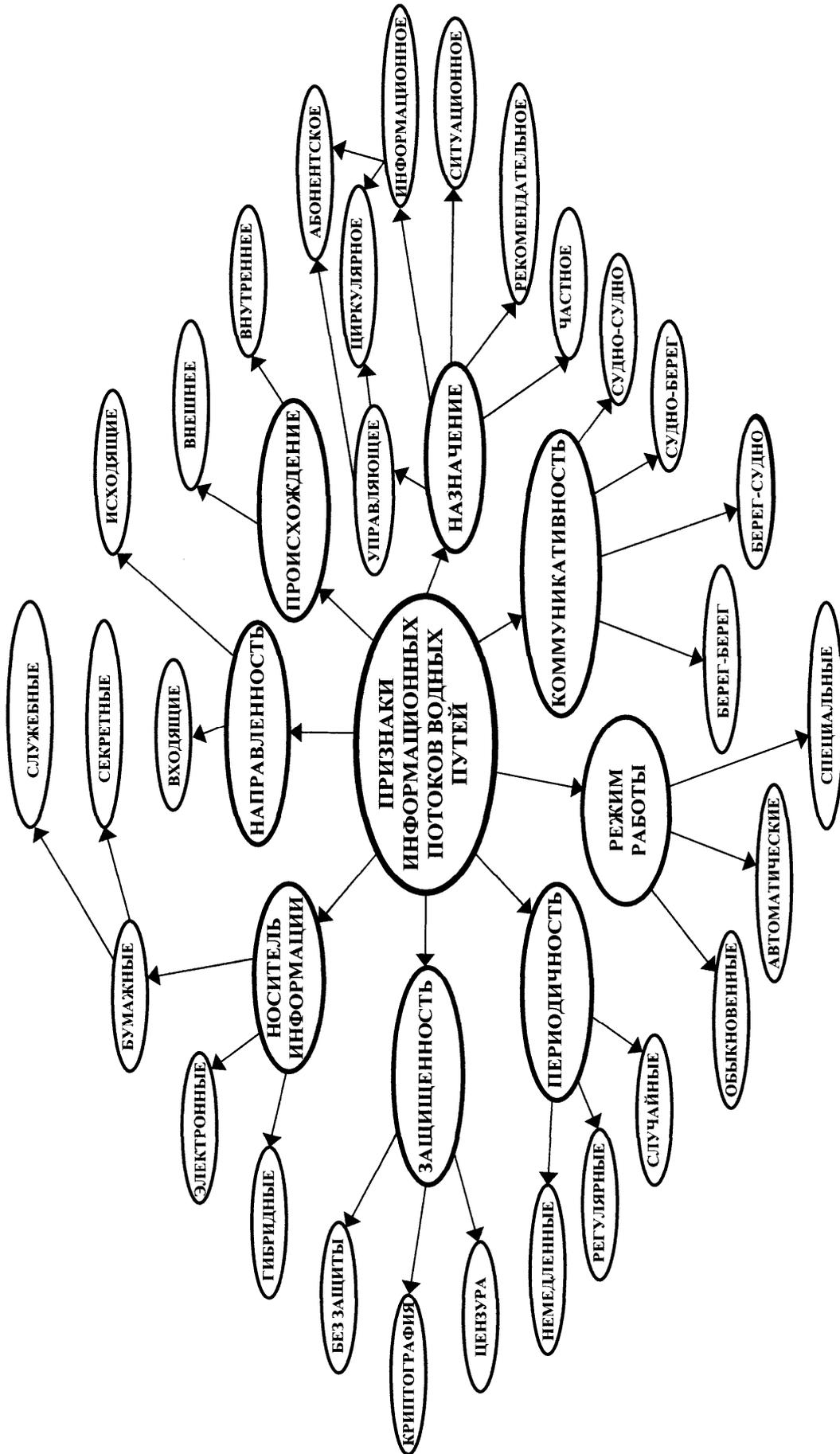


Рисунок 1 – Кластеризация информационных потоков водных путей по признакам

В качестве выборки используется множество категорий морских сообщений, каждая из которых представляется множеством определённого количества подклассов и представляется выражением

$$X^L = \{X_1(K_{n1}), X_2(K_{n2}), X_3(K_{n3}), \dots, X_L(K_{nm})\}, \quad (1)$$

где  $X_1(K_n)$  – множество классов морских сообщений;  $K_n$  – множество подклассов каждого из морских сообщений.

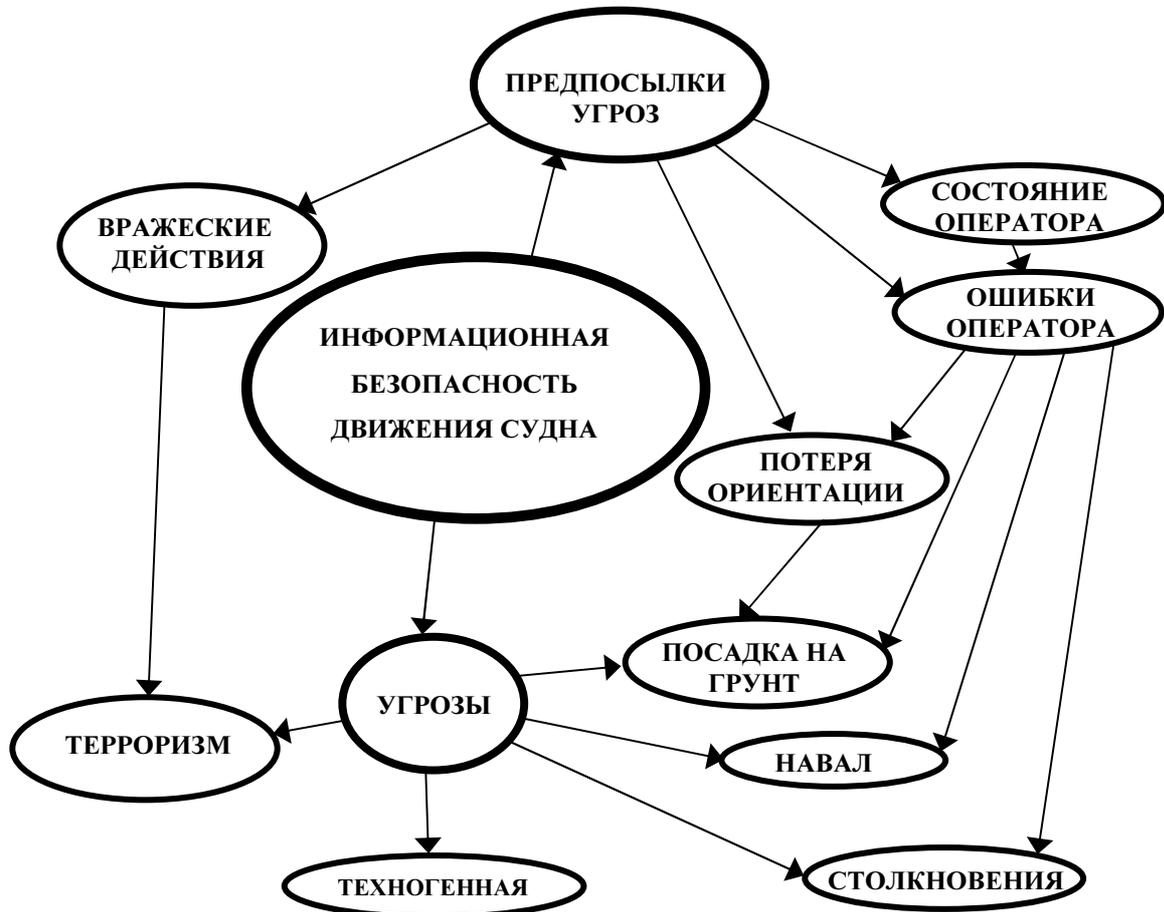


Рисунок 2 – Кластер информационной безопасности движения судна

При этом, в процессе кластеризации важным аспектом является понятие расстояния  $\rho(x, x')$ , которое при кластеризации играет роль количества объектов и связей, задействованных при формировании и передаче определённого потока морских данных.

Алгоритмизация формирования кластеров информационных потоков водных путей выполняется с использованием: эвристических графовых алгоритмов; статистических алгоритмов; алгоритмов иерархической кластеризации. Расчет кластерной структуры информационной безопасности целесообразно выполнять в симбиозе статистических методов с методом иерархической кластеризации. Достоинство метода иерархической кластеризации состоит в разделении потоков на непересекающиеся подклассы и построении с использованием графов дендрограмм, которые четко определяют внутри кластерные связи.

Для пространства информационной безопасности на водном пути вводится метрика, показывающая величину определения расстояний между кластерами. Такое пространство называется метрическим и для расчетов в нём кластеров информационной безопасности наиболее подходящей является метрика Евклида [10], в которой кластерные расстояния соответствуют выражению (2).

$$d_{ij} = \sqrt{\sum_{k=1}^n \frac{(x_{ik} - x_{jk})^2}{\sigma_k^2}}. \quad (2)$$

Данную метрику можно использовать для расчета внутрикластерных расстояний отдельных подгрупп, например, при расчете расстояний внутри кластера информационной безопасности водных путей или при определении объектов в вид подкластера («Угрозы» или «Предпосылки к угрозам»), так как объекты здесь имеют одинаковые характеристики, отличающиеся только интенсивностью отрицательного действия на движение судна. Рассмотрим кластер информационной безопасности движения судна и фрагмент исследования зависимости угроз на маршруте от внешних предпосылок, в условиях, когда объект постоянно находится в динамике (табл. 1). Для переменных, измеряемых в различных единицах или значительно различающихся по величине, при этом дисперсии по характеристикам отличаются друг от друга, метрика соответствует выражению (2). Приведём расчёт кластера информационной безопасности управления судном (рис.2.) с учётом фрагмента исследования зависимости угроз при движении судна по маршруту от внешних предпосылок, в условиях которых постоянно находится объект.

Расчет параметров безопасной судовой зоны осуществляется решением уравнения Радона-Никодима, как показано в работе [11]. Его решение в общем случае дает множество областей, из которых выбирается та, которая имеет минимальные размеры. При этом для поиска той же области используется нормальное распределение Гаусса, в котором плотность распределения выражается аналитическим выражением

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\left(\frac{(x - \mu_x)^2}{2\sigma_x^2} + \frac{(y - \mu_y)^2}{2\sigma_y^2}\right)\right], \quad (3)$$

где  $\mu_x, \mu_y$  – математические ожидания, соответственно, предпосылки «Потеря ориентации в навигационной обстановке» и угрозы «Посадка на грунт»,  $\sigma_x, \sigma_y$  – их дисперсии. На основании анализа выборки, составленной из количественных показателей аварийности за последние 20 лет, при возникновении предпосылки «Потеря ориентации» по факту возникновения угрозы «Посадка на грунт» находятся их математические ожидания и дисперсии по формулам

$$\sigma_x = \sqrt{\sum_{i=1}^5 p_{1i} (X_i - \mu_x)^2};$$

$$\mu_y = \sum_{i=1}^5 p_{2i} \cdot Y_i;$$

$$\mu_x = \sum_{i=1}^5 p_{1i} \cdot X_i;$$

$$\sigma_y = \sqrt{\sum_{i=1}^5 p_{2i} (Y_i - \mu_y)^2}$$

Предпосылка «Потеря ориентации»

<b>N<sub>1</sub></b>	0	1	2	3	4
<b>X</b>	13	8	4	0	0
<b>P<sub>1</sub></b>	0,52	0,32	0,16	0	0

Угроза «Посадка на грунт»

<b>N<sub>2</sub></b>	0	1	2	3	4
<b>Y</b>	14	9	4	0	0
<b>P<sub>2</sub></b>	0,52	0,33	0,15	0	0

В соответствии с полученными результатами, вероятность  $P_{ij}$  возникновения угрозы по данной предпосылке определяется выражением

$$P_{ij} = P_{1i} \cdot P_{2j}. \quad (4)$$

И, соответственно, для полученных вероятностей имеем следующие значения вероятностей возникновения угроз по данным предпосылкам.

$$P_{11} = 0,52 \cdot 0,52 = 0,2704$$

$$P_{22} = 0,32 \cdot 0,33 = 0,1056$$

$$P_{33} = 0,16 \cdot 0,15 = 0,024$$

Усредним полученный результат и получим вероятность  $P_m$  возникновения угрозы «Посадка на грунт» по факту возникновения предпосылки «Потеря ориентации в навигационной обстановке».

$$P_m = \frac{P_1 + P_2 + P_3}{3} = \frac{0,2704 + 0,1056 + 0,024}{3} = 0,4. \quad (5)$$

Произведем логарифмирование полученного результата и определим значение коэффициента расстояний в кластере  $A_{mn}$  по формуле

$$A_m = 10 |\ln P_m| = 9,00. \quad (6)$$

Коэффициенты расстояний по другим предпосылкам и соответствующим им угрозам проведены аналогичным методом и занесены в табл. 1.

Таблица 1 – Кластер информационной безопасности движения судна

Угрозы	Погодные условия	Состояние оператора	Ошибки оператора	Потеря ориентации	Терроризм
Посадка на грунт	8.00	8.00	7.25	9.00	7.25
Навал	6.25	7.00	6.00	7.00	6.00
Столкновение	7.50	6.25	7.00	8.00	5.00
Техногенная	4.00	3.00	1.25	1.50	2.00
Терроризм	2.00	1.00	1.00	0.50	1.50

При этом предлагается количественную оценку предпосылок рассчитывать по единой десятибалльной шкале. Такая квалиметрия лучше отображает влияние факторов на причины возникновения угроз. При расчете параметров кластеров каждый раз необходимо учитывать влияние факторов погодных условий - видимость, скорость ветра, высота волн, скорость течения и получать единый коэффициент для каждой угрозы. С использованием метрики Евклида первой категории получают симметричную матрицу расстояний (табл. 2),

Таблица 2 – Симметричная матрица расстояний кластера информационной безопасности

	Посадка на грунт (1)	Навал (2)	Столкновение (3)	Техногенная (4)	Терроризм (5)
Посадка на грунт (1)	0.00	8.05	7.2	2.06	3.20
Навал(2)	8.05	0.00	8.25	4.05	1.25
Столкновение (3)	7.03	8.25	0.00	5.25	7.80
Техногенная(4)	2.00	4.00	5.25	0.00	0.50
Терроризм (5)	3.00	1.00	8.10	0.50	0.00

которая является основой в кластерном анализе информационной безопасности движения судна. Метод иерархической кластеризации позволяет разделить потоки на непересекающиеся подклассы и строить с использованием графовых методов дендрограммы, которые четко определяют информационные кластерные связи. Результат применения иерархического метода к полученной матрице при использовании пакета статистического анализа, например STATISTICA, даёт информационную дендрограмму (рис. 3), которая показывает структуру кластера информационной безопасности движения судна (рис. 4).

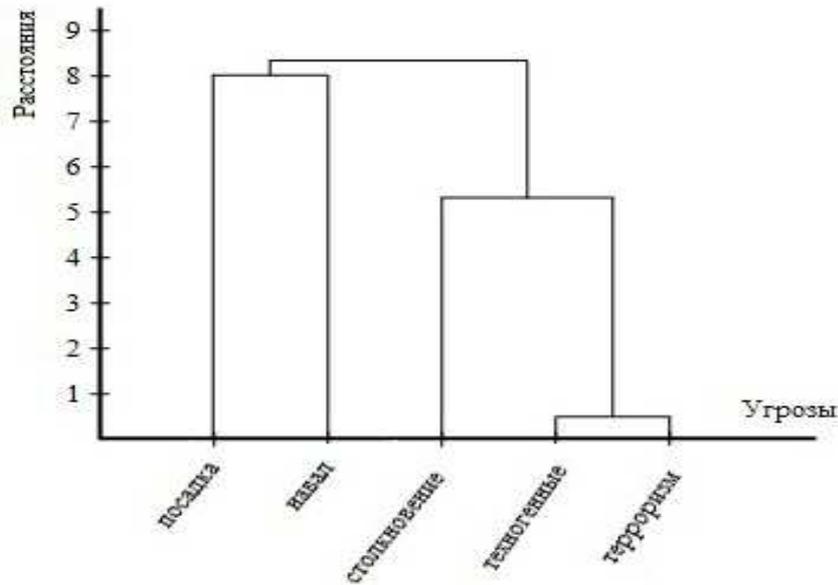


Рисунок 3 – Информационная дендрограмма упрощённого кластера безопасности движения судна

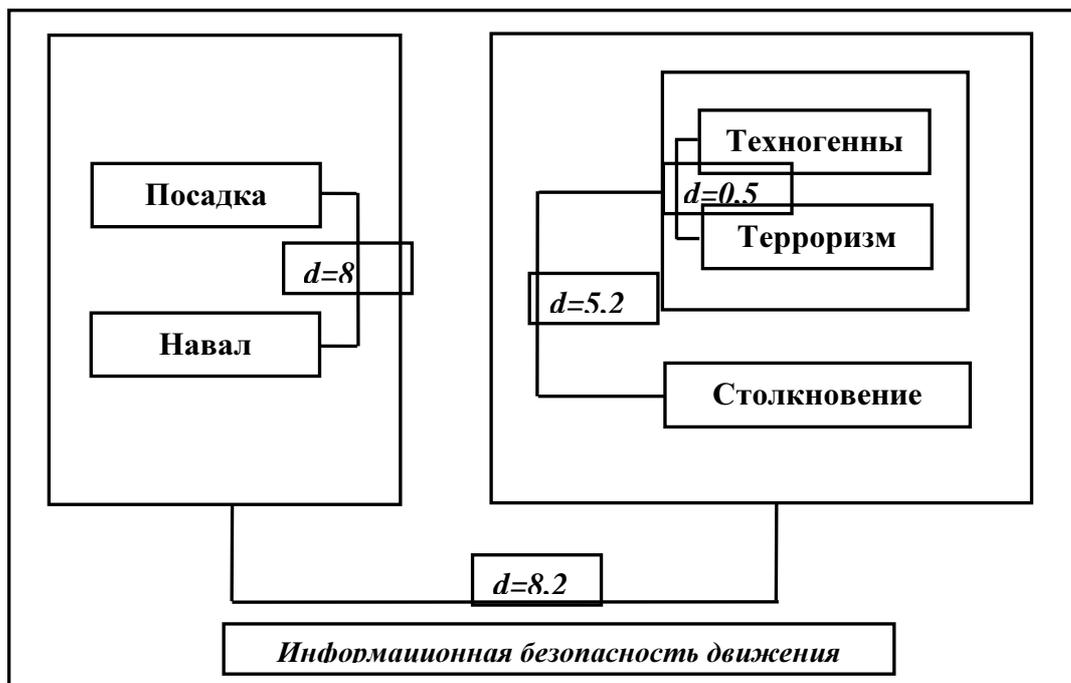


Рисунок 4 – Структура кластера информационной безопасности движения судна

На дендрограмме по горизонтальной оси отложены на одинаковом расстоянии номера объектов (угрозы), а по вертикальной оси – расстояния внутри кластеров. Полученная дендрограмма позволяет выявить близкие по параметрам объекты, входящие в кластер информационной безопасности, а также структуру самого кластера в данной точке или на заданном участке водного пути.

Видим, что сам кластер информационной безопасности состоит из своеобразных подкластеров, в которые включены объекты, близкие по параметрам и набору входных данных, что говорит о их равноценности

возникновения. Статистическая характеристика вероятности реализации события, которая включена в параметр расстояния, говорит о вероятности реализации самого события, а точнее совокупности событий на данном участке водного пути.

Таким образом, разбив объекты на классы, имеем классификацию, которая показывает, что для данного набора параметров предпосылок и угроз наиболее вероятными являются те аварийные ситуации, расстояние между которыми максимально. При этом выделяются пары объектов с расстояниями, которые, с точки зрения формул Эйлера, характеризуют равную вероятность аварийных событий.

**Выводы и предложения:**

1. Предложен подход к кластерному описанию морских сообщений для судовождения.
2. Приведено кластерное представление признаков информационных потоков на водных путях и реализован алгоритм симметричной матрицы расстояний кластера информационной безопасности движения судна.
3. Кластерный анализ морских сообщений эффективен в информационном обеспечении систем управления движением судов. Его использование позволяет классифицировать группы неоднородных потоков сведений и упростить решение задач управления судном.

**СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ**

1. Вильский Г. Б., Мальцев А. С., Бездольный В. В., Гончаров Е. И. Навигационная безопасность при лоцманской проводке судов : учебное пособие. / Под ред. А. С. Мальцева, Г. Б. Вильского. – Одесса – Николаев : ЧП «Феникс», 2007. – 456 с.

2. Шемякин О. М. Сучасні проблеми правової класифікації морських просторів // Судовождение : сб. научн. трудов / ОНМА. – Вып. 19. – Одесса : «ИздатИнформ», 2010. – С. 202-208.

3. Инфимовский С. Ю. Представление навигационной информации в виде карты поля ошибок обсерваций спутникового навигационного поля морской поверхности // Судовождение : сб. научн. трудов / ОНМА. – Вып. 19. – Одесса : «ИздатИнформ», 2010. – С. 76-80.

4. Головкин В. И. Состояние и перспективы развития систем управления движением судов в Российской Федерации // Сборник докладов 7-го международного научно-технического семинара «Состояние и перспективы развития береговых систем управления движением судов». – Николаев : Издание ГП «Дельта-лоцман» , 2004. – С. 5-12.

5. Вагущенко Л. Л., Вагущенко А. Л. Повышение информативности судовых систем предупреждения столкновений // Судовождение : сб. научн. трудов / ОНМА. – Вып.16. – Одесса : «ИздатИнформ», 2009. – С.18-26.

6. Бузовский Д. А., Алексишин В. Г. Определение вероятности получения высокоточной информации от обращенной радиолокационной

системы // Судовождение : сб. научн. трудов / ОНМА, Вып. 19. – Одесса : «ИздаИнформ», 2010 – С. 19-23.

7. Пятаков Э. Н., Алексишин В. Г. Определение угрозы ситуационного возмущения при опасном сближении судов // Судовождение : сб. научн. трудов / ОНМА, Вып. 19. – Одесса : ИздаИнформ, 2011 – С. 166-170.

8. Мамонтов В. В. Ранжирование и иерархия информации в системе управления движением судна. // Судовождение : сб. научн. трудов / ОНМА, Вып. 20. – Одесса : «ИздаИнформ», 2010. – С. 117-129.

9. Vilsky G. B., Nadich M. M. Information space research in Navigation. Proceedings of NAU– 2011 – №3 (48). – Pg. 95-102.

10. Вятчин Д. А. Нечёткие методы автоматической классификации. – Минск : Технопринт, 2004. – 219 с.

11. Вильский Г. Б., Надыч М. М. Теоретическое моделирование информационной безопасности водных путей // Судовождение : сб. научн. трудов / ОНМА, Вып.18. – Одесса : «ИздаИнформ», 2010. – С.48-57.

**Вильський Г.Б. КЛАСТЕРИЗАЦІЯ МОРСЬКИХ ПОВІДОМЛЕНЬ**

*У статті розглядаються проблеми кластерного аналізу морських повідомлень, приведено класифікацію методів кластеризації та на їх основі розроблені кластери потоків даних на водяних шляхах та інформаційної безпеки руху суден.*

*Ключові слова: кластерний аналіз, потоки даних, інформаційна безпека руху суден.*

**Вильський Г.Б. КЛАСТЕРИЗАЦІЯ МОРСЬКИХ ПОВІДОМЛЕНЬ**

*У статті розглядаються проблеми кластерного аналізу морських повідомлень, приведено класифікацію методів кластеризації та на їх основі розроблені кластери потоків даних на водяних шляхах та інформаційної безпеки руху суден.*

*Ключові слова: кластерний аналіз, потоки даних, інформаційна безпека руху суден.*