УДК 004.056.53

Метод комплексированной обработки информации для достижения достоверности данных в цифровых сенсорных системах

A method of integrated information processing to achieve data reliability in digital sensor systems

Карманова / Karmanova N.
Наталия Андреевна
(karmanova.ifmo@gmail.com)
ФГАОУ ВО «Национальный исследовательский Университет ИТМО»,
ассистент факультета безопасности информационных технологий.
г. Санкт-Петербург

Ключевые слова: комплексирование данных – data compilation; достоверность информации – information reliability; информационные потоки – information flows; вероятностные модели – probabilistic models; согласованность данных – data consistency; надёжность источников – source reliability; ложная информация – false information; вредоносная информация – malicious information; фильтрация данных – data filtering.

Статья описывает метод комплексирования данных для повышения достоверности информации при избыточности и разнородности источников. Используется вероятностная модель с учетом надёжности, согласованности и характеристик данных, например временных меток. Вводится динамическая адаптация весов источников, перекрёстная верификация сообщений и выявление ложной и вредоносной информации для достижения достоверности данных в цифровых сенсорных системах.

The paper describes a data aggregation method to improve the reliability of information in the presence of redundancy and heterogeneity of sources. A probabilistic model is used taking into account the reliability, consistency, and characteristics of data, such as timestamps. Dynamic adaptation of source weights, cross-verification of messages, and detection of false and malicious information are introduced to achieve data reliability in digital sensor systems.

Введение

Современные системы обработки данных и обеспечения информационной безопасности часто сосредоточены на анализе информации уже после ее поступления в систему, что приводит к недостаточному вниманию к этапу предварительной оценки информационного потока. Это создает риск проникновения в систему недостоверной, ложной или вредоносной информации, способной нарушить работу системы, снизить достоверность её решений или внести деструктивное воздействие на связанные процессы. Проблема усугубляется в условиях увеличения объемов и скорости поступающих данных, что затрудняет их своевременную классификацию и фильтрацию. Вследствие этого традиционные подходы, ориентированные на постфактум-анализ, демонстрируют низкую эффективность в предотвращении проникновения деструктивной информации. Непроведение предварительной проверки информационных потоков до их попадания в систему является недостатком, который подчёркивается в ряде исследований, рассматривающих распространение ложной или вредоносной информации. В следующих примерах исследований можно отметить подобные упущения: [1–5]. Эти исследования служат примерами, где методы анализа деструктивной информации применяются уже на стадии её распространения, а этап предварительной проверки или фильтрации потока данных до его попадания в системы либо отсутствует, либо недостаточно разработан. Это подчёркивает важность разработки решений, направленных на предотвращение интеграции вредоносной информации ещё на этапе предобработки данных.

Анализ проведенных в последние годы исследований [6–13] доказал справедливость утверждения о том, что реализация процедур определения ложной и вредоносной информации о сложных объектах и их состояниях в общем потоке поступающих данных представляет собой комплексную проблему и в каждом конкретном случае требует построения специальной системы защиты, обеспечивающей достоверность информации любого вида и области применения.

Представляется, что ключевым направлением борьбы с дезинформацией должно стать не только её выявление на отдельных этапах, но и обеспечение методик, которые напрямую повышают достовер-

ность итоговой информации [14–19]. Целью работы является выявление и противодействие распространению ложной и вредоносной информации.

Увеличение количества каналов передачи информации и их разнородность представляет собой как вызов, так и уникальную возможность. Использование разнородных данных от n источников позволяет применить метод их комплексирования, целью которого становится нивелирование ошибок, устранение противоречий и фильтрация ложных или вредоносных сигналов.

Метод комплексирования данных можно описать как процесс объединения информации от n различных источников или каналов с целью создания единого информационного потока повышенной достоверности. Такой метод позволяет снизить влияние ошибок, выбросов и ложных данных за счет перекрестной верификации данных, поступающих из разных источников.

1. Формализация задачи

Каждый источник S_i, где $i \in \{1,2,\dots,n\}$, генерирует поток данных D_i, представленный как множество сообщений:

$$D_i = \{d_{i1},\ d_{i2},\ \dots,\ d_{im_i}\}, \tag{1}$$

где m_i – количество сообщений от i-го источника.

Объединённый поток данных D_C от n источников формируется как их объединение:

$$D_C = \bigcup_{i=1}^{n} D_i = \{d_1, d_2, ..., d_m\},\ m \le \sum_{i=1}^{n} m_i . \tag{2}$$

D_C может содержать достоверные (согласованные между большинством источников), ложные (содержащие погрешности или противоречащие достоверным данным), вредоносные (намеренно искажённые) данные.

Примечание: стоит отметить, что объединенный информационный поток D_C может включать в себя данные, которые являются одновременно и ложными, и вредоносными. Одним из примеров может служить дублирующаяся информация, которая увеличивает время обработки общего массива данных. Это, в свою очередь, способно поставить под угрозу своевременность принятия ключевых решений, особенно в ситуациях, требующих высокой степени оперативности [20]. Однако разработанный алгоритм комплексированной обработки данных (рис. 1) включает в себя этап предварительной фильтрации дублей в блоке формирования объединенного потока данных D_{total}, направленный на идентификацию и исключение подобных нежелательных элементов. В рамках данной работы мы ограничиваем анализ до трёх вероятных категорий сообщения: ложное, вредоносное и достоверное. Такой подход позволяет с большей точностью определять характер информации и своевременно исключать потенциальные угрозы, сохраняя высокую точность, надёжность и эффективность функционирования системы даже в условиях ограниченных временных и вычислительных ресурсов. При том, что осуществляется комплексирование данных, полученных средствами добывания информации, построенных на различных физических принципах.

Каждое сообщение $d \in D_C$ из совокупного потока имеет три вероятности:

– достоверности $P(T|d)$,

– ложности $P(F|d)$,

– вредоносности $P(M|d)$.

Эти вероятности взаимодополняемы и нормализуются следующим образом:

$$P(T \mid d) + P(F \mid d) + P(M \mid d) = 1, \qquad d \in D_C . \tag{3}$$

Вероятности $P(T|d)$, $P(F|d)$, $P(M|d)$ зависят от:

– надёжности источников, из которых произошло сообщение d,

– степени согласованности данных d с другими источниками,

– метаданных (например, временных меток, тематического содержания и т. д.).

Задача состоит в том, чтобы в совокупном потоке D_C выявить данные, которые с высокой вероятностью являются искажёнными, ложными или вредоносными.

2. Вычисление надёжности источников

Каждый источник S_i характеризуется своей надёжностью Q_i, которая определяется как вероятностная мера того, насколько данный источник предоставляет достоверные данные. Q_i варьируется в пределах от 0 до 1:

$$Q_i \in [0,1]. \tag{4}$$

Надёжность Q_i можно вычислить на основе доступной информации о предыдущей достоверности источника:

$$Q_i = \frac{\text{число достоверных данных из } S_i}{\text{общее число данных из } S_i}. \tag{5}$$

Альтернативно Q_i можно оценить на основе машинного обучения, исторической референции данных и уровня конфликта сообщений, связанного с этим источником (описано далее).

3. Комплексирование потоков данных

Подход «взвешенного комбинирования вероятностей» строится на предположении, что сообщения, поступающие от источников, имеют разную степень достоверности, которая определяется за счёт надёжности источников (Q_i) и контекстуальной значимости сообщения. Этот метод позволяет учесть множественные аспекты данных через механизм вероятностного анализа и взвешивания с использованием индивидуальных характеристик каждого источника.

Взвешенное комбинирование вероятностей подразумевает вычисление объединённой вероятности достоверности $P(T|d)$, ложности $P(F|d)$ и вредоносности $P(M|d)$ каждого сообщения $d \in D_C$ на основе вероятностей, рассчитанных для этого сообщения разными источниками ($S_1, S_2, ..., S_n$).

Для каждого сообщения $d \in D_C$ взвешенные вероятности различных характеристик сообщения $P(T|d)$, $P(F|d), P(M|d)$ вычисляются как:

$$P(X|d) = \frac{\sum_{i=1}^{n} W_i * P_i(X|d)}{\sum_{i=1}^{n} W_i}, \quad (6)$$

где:

$X \in \{T, F, M\}$ – одна из характеристик сообщения (достоверность, ложность или вредоносность);

$P_i(X|d)$ – предварительная вероятность, рассчитанная источником S_i для сообщения d на основе анализа данных;

W_i – вес источника S_i, определяющий вклад данного источника в итоговую вероятность сообщения, $W_i = Q_i$.

$\sum_{i=1}^{n} W_i$ – нормировочный коэффициент, обеспечивающий корректность вероятностной интерпретации.

В результате объединённые вероятности сообщения рассчитываются следующим образом для каждой из характеристик (T, F, M):

1. Вероятность достоверности $P(T|d)$:

$$P(T|d) = \frac{\sum_{i=1}^{n} Q_i * P(T|d)}{\sum_{j=1}^{n} Q_i}. \quad (7)$$

2. Вероятность ложности $P(F|d)$:

$$P(F|d) = \frac{\sum_{i=1}^{n} Q_i * P(F|d)}{\sum_{j=1}^{n} Q_i}. \quad (8)$$

3. Вероятность вредоносности $P(M|d)$:

$$P(M|d) = \frac{\sum_{i=1}^{n} Q_i * P(M|d)}{\sum_{j=1}^{n} Q_i}. \quad (9)$$

На основании этих вероятностей сообщение d классифицируется по категориям.

Учет согласованности позволяет выявить сообщения, подтвержденные несколькими независимыми источниками, что автоматически повышает вероятность их достоверности, а также исключить или снизить влияние противоречивых данных.

Согласованность данных играет ключевую роль в задачах классификации и фильтрации. Она позволяет учитывать коллективное подтверждение информации для её достоверности, минимизирует влияние шумов, уменьшает противоречия и помогает адаптивно перераспределять веса источников. В данном разделе рассматривается математическое описание подхода к учету согласованности данных.

Для формального анализа согласованности вводится коэффициент согласованности $k(d)$ для конкретного сообщения d, представляющий собой меру коллективного подтверждения или опровержения данных, предоставленных источниками. Этот коэффициент рассчитывается как доля источников, которые поддерживают сообщение d, по отношению к общему числу источников:

$$K(d) = \frac{\sum_{j=1}^{n} Q_i * C_i(d)}{\sum_{i=1}^{n} Q_i}, \quad (10)$$

где:

$C_i(d)$ – индикатор согласованности, показывающий, поддерживает ли источник S_i сообщение d. $C_i(d) = 1$, если источник S_i подтверждает сообщение d, $C_i(d) = 0$, если источник S_i не подтверждает сообщение d.

Q_i – надежность источника S_i.

$\sum_{i=1}^{n} Q_i$ – суммарный вес всех источников.

Коэффициент $K(d)$ принимает значения от 0 до 1:

– $K(d) \to 1$: сообщение подтверждается большинством надежных источников, и вероятность его достоверности увеличивается.

– $K(d) \to 0$: сообщение либо не подтверждается большинством источников, либо противоречит более надежным источникам, что снижает вероятность его достоверности.

Этот коэффициент может быть расширен дополнительными параметрами, учитывающими временные метки сообщений или другие метаданные, если это важно для конкретной задачи.

На этапе классификации данных каждое сообщение из объединённого потока данных D_C классифицируется в одну из трёх категорий: достоверные данные (T), ложные данные (F) и вредоносные данные (M). Механизм классификации базируется на вероятностных характеристиках, рассчитанных ранее (взятых из этапов комплексирования, включая взвешенное комбинирование вероятностей и анализ согласованности), а также на применении пороговых значений, искомых алгоритмическим или эмпирическим путём.

Для классификации сообщений используется принцип максимальной вероятности и задание пороговых значений θ для каждой из категорий. Пороговые значения θ в данном контексте представляет собой фиксированные параметры, определяющие границы для отнесения сообщения к определённой категории (ложное, вредоносное или достоверное). Однако важно отметить, что θ является независимой величиной, и она не связана напрямую с нормализацией вероятностей классификации.

Стратегии выбора порогов:

1. Эмпирическая настройка: пороговые значения задаются экспертно на основе априорных данных или предыдущего опыта работы с системой. Для большинства задач рекомендуются следующие приближённые пороги:

– $\theta_T = 0{,}7$ (достоверные данные),

– $\theta_F = 0{,}4$ (ложные данные),

– $\theta_M = 0{,}2$ (вредоносные данные).

2. Метрики производительности: для оптимизации порогов используются показатели, оценивающие точность предсказаний и степень охвата истинных значений модели, а также их сбалансированность. На основе исторических данных определяется пороговое значение, минимизирующее вероятность ошибок первого и второго рода.

3. Адаптивная настройка: значения θ_T, θ_F, θ_M автоматически настраиваются в процессе работы модели с использованием алгоритмических методов, таких как градиентный спуск или байесовская оптимизация. Это обеспечивает повышение точности модели и уменьшение ошибок благодаря динамической подстройке под характеристики данных. Следует отметить, что исследование методов адаптивной настройки параметров само по себе представляет отдельное направление в области машинного обучения и оптимизации, которое заслуживает более глубокого анализа и может стать темой для самостоятельной научной статьи.

4. Задача мультиклассовой классификации: если вероятности категорий $P(T|d), P(F|d), P(M|d)$ распределены без чёткой доминанты, возможно задание промежуточных состояний для обработки сложных данных, которые нуждаются в ручной доработке (например, сообщения с вероятностями [0,4,0,3,0,3])

Каждое сообщение d классифицируется следующим образом:

– $d \in T$, если $P(T|d) \geq \theta_T$;

– $d \in F$, если $P(F|d) \geq \theta_F$;

– $d \in M$, если $P(M|d) \geq \theta_M$.

При этом выполняются условия нормализации вероятностей (3).

Таким образом, каждое сообщение однозначно попадает в одну из трёх категорий в зависимости от его вероятностной характеристики. В случае, когда вероятности не достигают установленных порогов, сообщение относится к информационному шуму, который удаляется на этапе фильтрации в блоке формирования достоверного потока данных D_{output} в разработанном алгоритме (рис. 1).

4. Повышение достоверности данных

Повышение достоверности данных является ключевой целью комплексирования информации из множества источников в рамках многостадийной обработки потоков данных. На этом этапе система стремится не только классифицировать данные и исключить ложные или вредоносные сообщения, но и активно улучшить качество и достоверность итогового потока данных. Это достигается за счёт следующих подходов:

– *Динамическая адаптация надёжности источников:*

Надёжность источников Q_i, рассчитанная на этапе 2, пересматривается на основе качества предоставленных данных. Этот процесс напоминает механизм обратной связи. Перерасчёт Q_i осуществляется с помощью следующей формулы:

$$Q_i^{new} = Q_i^{current} + \lambda \cdot \left(\frac{N_T}{N_D} - Q_i^{current} \right), \tag{11}$$

где:

· Q_i^{new} – обновлённая надёжность источника;

· $Q_i^{current}$ – текущая надёжность источника;

· N_T – количество достоверных сообщений от источника;

· N_D – общее количество сообщений от источника;

· λ – параметр скорости адаптации (обычно λ [0.1,0.5]).

Этот процесс обеспечивает то, что достоверные источники будут увеличивать свою надёжность, а ненадёжные – снижать свой вес, что минимизирует их влияние на итоговые данные.

– *Учет согласованности в комбинировании данных:*

Согласованность, обсуждаемая в п. 3, играет критическую роль в повышении достоверности. Принимается предположение, что сообщения, возвращающие высокую степень согласованности $K(d)$, заслуживают большего доверия. Итоговая вероятность достоверности сообщения корректируется с учётом согласованности следующим образом:

$$P'(T|d) = P(T|d) \cdot \left(1 + K(d)\right), \tag{12}$$

где $K(d)$ – коэффициент согласованности сообщения, рассчитываемый на основе подтверждения несколькими источниками (см. п. 3). Увеличение вероятности $P'(T/d)$ для сообщений, поддержанных надёжными источниками, повышает общее качество выхода системы.

– *Снижение влияния ложных данных:*

Для исключения ложных данных из итогового потока используются как пороговые значения (θ_T и θ_M), так и исторический анализ. Сообщения, которые классифицируются как ложные (F) или противоречат консенсусу большинства источников, обладающие низким $P(T/d)$, удаляются из итогового потока.

– *Детектирование вредоносных данных:*

Повышение достоверности данных включает не только фильтрацию, но и активное выявление вредоносных данных. Для этого вводятся промежуточные метрики, такие как:

Коэффициент конфликтности сообщений $C_{conflict}(d)$:

$$C_{conflict}(d) = \frac{1}{|S|} \cdot \sum_{i=1}^{n} \left(1 - C_i(d)\right) \cdot Q_i, \tag{13}$$

где $C_i(d)$ отражает согласованность источника S_i по отношению к сообщению d.

Сообщение помечается как вредоносное и удаляется из итогового потока, если его коэффициент конфликтности $C_{conflict}(d)$ превышает установленный порог τ. Значение τ выбирается эмпирически, например, τ=0,6.

– *Использование метаданных для дополнительной проверки:*

Метаданные, такие как временные метки, географические данные или размер сообщений, также могут быть источником для анализа достоверности. Например:

– Временная согласованность: если поток сообщений от одного источника согласован по времени с потоком от других, его сообщения получают дополнительный вес.

– Аномалии в структуре данных: сообщения, сильно отличающиеся по объёму или содержанию от медианы, требуют повторного анализа.

Следуя всем описанным выше этапам и процедурам, формируется финальный поток данных D_{output}, состоящий исключительно из сообщений, прошедших проверку на достоверность: $D_{output}=D_T$. Поток D_{output} интегрируется в систему и используется для принятия решений или передачи конечному пользователю.

Этап повышения достоверности направлен на улучшение общего качества выходного потока данных за счёт пересмотра надёжности источников, учёта согласованности сообщений, удаления ложной и вредоносной информации и активной адаптации весов в системе. Перечисленные методы повышения достоверности делают систему более устойчивой, гибкой и способной эффективно работать даже в условиях высокошумной информации.

Результаты экспериментальных исследований по оценке формирования и предоставления достоверных данных

На основе написанной программы, реализующей алгоритм комплексированной обработки данных, были проведены экспериментальные исследования для оценки формирования и предоставления достоверных данных.

Блок-схема алгоритма комплексированной обработки данных предоставлена на рис. 1.

Алгоритм направлен не только на выявление достоверных данных, но и на детектирование ложных и вредоносных потоков, что минимизирует риски использования искажённой информации. Новизна предлагаемого алгоритма заключается в его способности одновременно учитывать множество факторов (надёжность, согласованность, конфликтность и подтвержденность), динамически адаптировать параметры в процессе работы и классифицировать данные по трём уровням. За счёт таких особенностей он существенно превосходит существующие аналоги по точности, устойчивости и универсальности, что подтверждается результатами экспериментальных исследований.

Целью экспериментальных исследований стало подтверждение гипотезы о том, что метод комплексирования данных от n источников способен существенно повысить достоверность формируемой информации за счёт использования избыточности и перекрёстной корреляции данных между источниками и исключения низкодостоверных источников и сообщений с информацией, выходящей за рамки согласованных данных. Результаты проведённых экспериментов позволили выявить эффективность предложенного метода и подтвердить его преимущества.

Для проверки гипотезы были смоделированы различные сценарии объединения данных от n источников, где источники имели разные уровни надёжности (Q_i). Данные могли быть достоверными (согласованными между большинством источников), ложными (содержащими погрешности или противоречащие достоверным данным), вредоносными (намеренно искажёнными).

Параметры оценивались при разном уровне избыточности данных (процент дублирующих или схожих сообщений в общей выборке).

Метрики оценки:

Достоверность объединённых данных (D): определялась как отношение количества достоверной информации в итоговом потоке D_{total} к общему объёму данных от всех источников:

$$D = \frac{\text{Количество достоверных сообщений}}{\text{Общее количество обработанных сообщений}}. \quad (14)$$

Согласованность (C): оценка доли данных, подтверждённых большинством источников (перекрёстных коррелированных данных):

$$C = \frac{\text{Количество подтверждённых данных}\left(\text{более 50\% голосов}\right)}{\text{Общее количество обработанных сообщений}}. \quad (15)$$

Доля исключённых ненадёжных сообщений (E): процент сообщений, которые были исключены из итогового потока данных из-за выявленных несоответствий или низкой надёжности источника:

$$E = \frac{\text{Количество исключенных сообщение}}{\text{Общее число обработанных сообщений}}. \quad (16)$$

В эксперименте участвовало 10 источников данных (n=10) с разной начальной надёжностью, случайным образом установленной в пределах от 0,3 до 0,9. Всего было сгенерировано 10,000 сообщений, из которых 60 % – достоверные данные (описание одного и того

Рис. 1. Блок-схема алгоритма комплексированной обработки данных

Таблица 1

Результаты использования избыточности данных и перекрёстной корреляции

Уровень избыточности (%)	*D* (Достоверность итогового потока)	*C* (Согласованность итогового потока)
10	0,75	0,83
30	0,85	0,89
50	0,9	0,93
70	0.95	0,97
90	0.98	0,97

Таблица 2

Результаты исключения недостоверных данных

Уровень шума в данных (%)	*E* (Исключённые данные)
10	0,12
30	0,22
50	0,36
70	0,51

же события от нескольких источников), 30 % – ложные данные (включая шумовые данные, незначительные расхождения), 10 % – вредоносные данные (полностью искажённые или противоречивые).

Избыточность данных (частота получения совпадающих данных от разных источников) варьировалась от 10 % до 90 %.

Этапы эксперимента:

1. Комбинация данных от всех источников: для каждого сообщения проверялось, предоставляло ли его несколько источников.

2. Перекрёстная корреляция: оценивалась согласованность сообщений (анализ временных меток, полей содержимого).

3. Исключение недостоверных сообщений: на основании расчёта надёжности источников (Q_i) и соответствия данных большинству источников.

4. Формирование итогового потока D_{total}.

5. Оценка метрик (*D,C,E*) для каждого уровня избыточности данных и уровня шума.

Повышение достоверности информации (*D*):

При увеличении избыточности (наличии перекрытия сообщений между источниками) достоверность объединённых данных значительно увеличилась:

– При низкой избыточности (10 %): $D \approx 0{,}75$.

– При высокой избыточности (90 %): $D \approx 0{,}98$.

Это объясняется тем, что при высокой степени избыточности система могла легче идентифицировать ошибки и вредоносные данные путём анализа перекрёстной корреляции между источниками. Полученные результаты зависимости достоверности информации от избыточности данных представлены в таблице 1.

Улучшение согласованности данных (C):

При использовании перекрёстной корреляции удалось достичь высокого уровня согласованности данных (таблица 1). Даже при низких значениях избыточности алгоритм обеспечивал высокие показатели согласованности. При средней и высокой степени избыточности (50 %–90 %) согласованность достигала $C \geq 0{,}93$.

Исключение недостоверных данных (E):

На этапе обработки исключалось до 98 % вредоносных данных. При этом 100 % дублирующих сообщений также удалялись как избыточные. Ошибочные сообщения от ненадёжных источников или конфликтующие данные составляли основную часть исключённых элементов (при средней надёжности источников). Результаты исключения недостоверных данных представлены в таблице 2.

Сравнение с некоррелированными данными (без комплексирования):

Для сравнения достоверность данных *D* (без корреляции), обработанных без перекрёстной корреляции, составила 0,65 при низкой избыточности данных и 0,81 при высокой избыточности.

Таким образом, метод комплексирования данных обеспечил рост достоверности итогового потока *D* на 15 %–20 % в сравнении.

Выводы

Методы, описанные в данной статье, позволяют не только улучшить качество обработки данных, но и развивать механизмы защиты от ложной или вредоносной информации в условиях информационной войны, кибератак или массовой дезинформации.

Преимущества предложенных методов повышения достоверности:

1. Гибкость и адаптивность:

– Методы динамической корректировки надёжности источников позволяют системе адаптироваться к изменяющимся условиям. Например, если новый источник со временем продемонстрирует высокую достоверность, он сможет получить более значительный вес в последующей обработке данных.

– Система адаптирует свои параметры (пороги, весовые коэффициенты и др.), чтобы реагировать на проявляющиеся типы ошибок обработки информации – упреждающий ответ на характерные схемы дезинформации.

2. Многослойность подхода:

В основу системы заложен многоуровневый анализ данных: от индивидуальных характеристик сообщений до глобальной оценки всей совокупности данных. Это делает подход более глубоким и проактивным за счёт перерасчёта показателей согласованности, вероятностей категорий данных и взаимодействия разных источников.

3. Обнаружение и блокировка вредоносной информации:

Одной из ключевых функций системы является активное выявление потенциально вредоносных сообщений и минимизация их влияния. Благодаря статистическим и вероятностным методам становится возможным предотвращение сбоев системы из-за некорректной или враждебной информации.

4. Устойчивость к информационным атакам:

Критически важным аспектом является защита данных от внешних атак. Применение описанных алгоритмов обеспечивает снижение вероятности успешной дезинформации или манипуляции данными на выходе системы.

5. Улучшение качества анализа на следующем уровне:

Итоговый поток данных, построенный на результатах комплексирования, классификации и повышения достоверности, представляет собой основу для дальнейших аналитических процессов. Эти данные являются чистыми (с минимальным количеством ложной информации), согласованными и корректными, что даёт возможность улучшить предсказательные модели, аналитические системы и процедуры принятия решений.

Литература

1. Fake news detector using deep learning / A. Akshansh, C. Aksh, R. Gaurav [et al.] // International Journal of Advanced Research. – 2024. – No. 11 (04). – P. 1612–1621.

2. Revisiting Fake News Detection: Towards Temporality-aware Evaluation by Leveraging Engagement Earliness / J. Kim, J. Lee, Y. In [et al.] // arXivLabs. – 2024. – No. 2411. – P. 1–11.

3. Zhou, X. A survey of fake news: Fundamental theories, detection methods, and opportunities / X. Zhou, R. Zafarani // ACM Computing Surveys. – 2020. – No. 53 (5). – P. 1–40.

4. Привалов, А. Н. Поиск фейковых сайтов с использованием метода определения визуального сходства страниц / А.Н. Привалов, В.А. Смирнов // Известия Тульского государственного университета. Технические науки. – 2022. – № 9. – С. 260–264.

5. Обеспечение целостности данных посредством частичной «фрагментации» данных / А.К. Куртов, Д.О. Комиссаров, Е.И. Жабов, А.О. Сорокин // Современные научные исследования и инновации. – 2023. – № 9. – [Электронный ресурс]. – URL : https://web.snauka.ru/issues/2023/09/100716 (дата обращения: 05.03.2025).

6. Hammouchi, H. Evidence-Aware Multilingual Fake News Detection / H. Hammouchi, M.M. Ghogho // IEEE Access. – 2022. – No. 99. – P. 1–11.

7. Zhou, Y. The Silent Saboteur: The Impact and Management of Malicious Word-Of-Mouth in The Digital Age / Y. Zhou // Highlights in Business Economics and Management. – 2024. – No. 41. – P. 381–386.

8. The impact of malicious nodes on the spreading of false information / Z. Ruan, B. Yu, X. Shu [et al.] // Chaos: An Interdisciplinary Journal of Nonlinear Science. – 2020. – No. 30 (8). – P. 083101.

9. Sneha, B. Detecting malicious Twitter bots using machine learning / B. Sneha // International Journal for Research in Applied Science and Engineering Technology. – 2024. – Vol. 12, Iss. 3. – P. 1457–1463.

10. Wesam, H. A. Opinion mining for fake recommendations in e-commerce: A machine learning approach using LightGBM / H.A. Wesam, A. Ragheed, H.A. Yossra // AIP Conference Proceedings. – 2025. – No. 3169, B. 030015. – P. 1–11.

11. Козлов, В. В. Обоснование облика системы защиты стартового комплекса от деструктивных воздействий / В.В. Козлов, А.В. Лагун, В.А. Харченко // Известия ТулГУ. Технические науки. – 2023. – № 1.– С.259–265.

12. Лубенцов, А. В. Синтез метода оценки эффективности системы информационной безопасности / А.В. Лубенцов // Известия вузов. Электроника. – 2024.– Т. 29, № 1. – С. 118–129.

13. Метод выявления и противодействия распространению вредоносной информации в роевых робототехнических системах в процессе распределения задач / А.С. Павлов, В.И. Петренко, Ф.Б. Тебуева [и др.] // Инженерный вестник Дона. – 2022. – № 11 (95). – С. 382–410.

14. Gagolewski, M. Data Fusion: Theory, Methods, and Applications / M. Gagolewski. – 2022. – URL : https://arxiv.org/abs/2208.01644 (дата обращения: 19.05.2025).

15. Yuqing, T. Research on Multi-sensor Data Fusion Technology / T. Yuqing, B. Jiujun, C. Xuebo // Journal of Physics: Conference Series. – 2020. – Vol. 1624. – P. 032046.

16. Galich, A. A data fusion approach for providing valid annual passenger transport statistics during the COVID-19 pandemic / A. Galich, C. Eisenmann, K. Kцhler // European Transport Research Review. – 2025. – Vol. 17 (1).

17. Blasch, E. Handbook of multisensor data fusion: theory and practice / E. Blasch. – Boca Raton : CRC Press, 2017. – 870 p.

18. Multi-source data fusion for cyberattack detection in power systems / A. Sahu, Z. Mao, P. Wlazlo [et al.] // arXiv preprint arXi. – 2021. – P. 1–18.

19. Remote sensing data fusion approach for estimating forest degradation: A case study of boreal forests damaged by Polygraphus proximus / S. Illarionova, P. Tregubova, I. Shukhratov [et al.] // Frontiers in Environmental Science. – 2024. – Vol. 12. – P. 1412870.

20. Грищенко, Л. Л. Применение инструментария ситуационных центров для обеспечения ситуационного управления / Л.Л. Грищенко, В.В. Данилов // Право и государство: теория и практика. – 2023. – № 7 (223). – С. 166–169.