

ОПРЕДЕЛЕНИЕ ЭНТРОПИИ КАК МЕРЫ ИНФОРМАЦИИ ПРИ СОПОСТАВЛЕНИИ ПРОГНОЗНЫХ И ФАКТИЧЕСКИХ ПОКАЗАТЕЛЕЙ ПРЕДПРИЯТИЯ

Дулесов А.С., Хрусталеv В.И.

Хакасский государственный университет им. Н.Ф. Катанова, Абакан, Россия, E-mail: khsukhsu@mail.ru

Рассматриваются возможности применения в задаче анализа прогнозных оценок современных информационных технологий. В числе имеющихся информационных технологий, на основе теории информации, предусмотрены методологии определения меры информации. Для этого в качестве исходных предпосылок выбрано несовпадение прогнозируемых и фактических данных о поведении системы. Среди существующих мер определения неопределенности информации принята энтропия. Статья включает в себя модель и алгоритм определения энтропии информации по Шеннону и В-энтропии, учитывающие геометрическое обобщение данных. Алгоритм позволяет выполнять предварительную обработку данных с целью получения разностей через сопоставление прогнозных и фактических данных. Для обработки большого количества данных предусмотрена процедура построения гистограммы случайных распределений. Получение конечного результата рассматривается на конкретном примере и с помощью разработанного программного обеспечения. Полученные значения энтропии могут служить критерием для анализа результатов прогноза.

Ключевые слова: энтропия; мера информации; прогнозирование.

THE CALCULATION OF THE ENTROPY AS AN INFORMATION MEASURE WHILE COMPARING THE PREDICTING AND FACTUAL DATA OF AN ENTERPRISE

Dulesov A.S., Khrustalev V.I.

Khakas State University named after N.F. Katanov, Abakan, Russia, E-mail: khsukhsu@mail.ru

The article deals with the application possibilities in the analysis of modern information technologies predicting assessment. The theory of information technologies includes a methodology of an information measure exposure. To do that we should choose the convergence of predicting and factual data about the system behavior.

An entropy is accepted as a measure for the calculation of the information uncertainty. The article includes the model and the algorithm of an entropy information calculation according to Shannon and B-entropy, which take into account the facts of Geometry. The algorithm allows carrying out the preliminary data handling in order to get the difference by comparing predicting and factual data. For a great amount of data handling the procedure of a casual distribution bar chart is provided. The process of the final result is illustrated by the example with the help of the software. The entropy's values can serve as a criterion for the prediction's results.

Key words: entropy, information measure, prediction.

Введение. Процесс прогнозирования включает в себя формирование системы количественных и качественных показателей развития предприятия, анализ поведения которой позволяет оценивать темпы, пропорции, тенденции его развития на перспективу. Прогнозирование является одной из основных функций планирования на предприятии и необходимым элементом эффективного управления. Важность прогнозирования возрастает ввиду того, что производство становится более сложным бизнес-процессом в

деятельности компаний. Её конкурентоспособность напрямую определяется эффективностью используемой системы планирования и управления производством.

При прогнозировании необходимо учитывать многие факторы:

- внешние: география и природные ресурсы; экология; экономические условия и конкуренция; факторы социальной и культурной сферы; влияние правительства и законодательства; техногенные факторы и др.;

- внутренние: неплатежи – финансовая неустойчивость; недостаток оборотных средств; состояние общепроизводственных фондов, технологий, производственных мощностей; кадровый потенциал – состояние, подготовка и переподготовка и др.

Указанные факторы обуславливают наличие риска (экономического, коммерческого и т.п.) и в разной степени оказывают влияние на систему прогнозирования, на дальнейшую подготовку и реализацию плана (или проектного решения). Отсутствие у разработчиков прогноза необходимой и достоверной информации о роли и степени воздействия факторов в конечном итоге приводит к тому, что фактические показатели существенно отличаются от прогнозных значений. В процессе прогнозирования необходим анализ показателей для последующей минимизации степени неопределенности как способа принятия обоснованных решений во всех аспектах планирования.

Чтобы минимизировать степень неопределенности в процессе реализации прогнозов, необходимо идентифицировать область потенциального риска, определить вероятность его возникновения и потенциальные последствия. Использование инструментов (например, определения меры неопределенности информации), осуществляющих оценку факторов риска, помогает менеджеру вырабатывать решения по выбору наиболее достоверного варианта прогноза.

Одним из критериев оценки меры неопределенности является энтропия, которая способна отразить хаотическое поведение объекта. Она представляет собой вещественно-значную функцию, зависящую от вероятности появления событий. Особенности применения энтропии рассматривались авторами в работах [1–3].

Далее предлагается метод определения энтропии как меры неопределенности информации в задаче сопоставления прогнозных (плановых) показателей с фактическими.

Модель определения энтропии. В деятельности предприятия за определенный период времени информация о прогнозируемых и фактических результатах работы является детерминированной. Однако отклонения фактических от прогнозных значений для фиксированных моментов времени несут в себе вероятностную природу флуктуаций. Разности этих значений (исходы) представляют собой множество $X = \{x_1, x_2, \dots, x_i, \dots, x_M\}$, состоящее из M элементов. На данном множестве должно быть задано распределение

вероятностей $p(x)$, если каждому исходу x_i поставлено в соответствие число $p(x_i)$ такое, что для всех $i = 1, 2, \dots, M$ должно быть $p(x_i) \geq 0$, $\sum p(x_i) = 1$. Множество X вместе с заданным на нём распределением вероятностей является дискретным вероятностным ансамблем или дискретным ансамблем $\{X, p(x_i)\}$.

Имея в своем распоряжении дискретный ансамбль с M возможными состояниями (разностями):

$$X = \left[\begin{array}{c} x_1, x_2, \dots, x_i, \dots, x_M \\ p_1, p_2, \dots, p_i, \dots, p_M \end{array} \right], p_i = p(x_i) \geq 0, \sum_{i=1}^M p_i = 1, \quad (1)$$

можно приступить к определению энтропии. При большой размерности ансамбля количество полученных вероятностей будет также большим. Чтобы избежать чрезмерной размерности данных, можно воспользоваться алгоритмом построения статистического распределения исходов по количественному признаку, то есть построением гистограммы. Тогда можно получить меньшее число вероятностей при незначительных потерях в точности расчетов.

Поскольку вероятности p_i не равны между собой, то естественным можно считать требование, чтобы мера неопределенности была непрерывной функцией вероятностей p_i , $i = 1, 2, \dots, M$ элементов дискретного ансамбля. Удовлетворяющая этому требованию мера предложена К. Шенноном и называется энтропией:

$$H(X) = -\sum_{i=1}^M p(x_i) \log p(x_i). \quad (2)$$

В данном вероятностном ансамбле энтропия $H(X)$ позволяет измерять информацию количественно, а также является априорной мерой неопределенности выбора. Применить выражение (2) для нашего случая не составляет особого труда.

По результатам $H(X)$ для различных состояний реализации плана можно судить о мере неопределенности выбора прогнозных вариантов. Если все вероятности p_i будут равны между собой (что маловероятно), $H(X)$ примет максимальное значение, в том числе и по формуле Хартли.

Представленная формула Шеннона позволяет находить значение энтропии только через вероятностные свойства отклонений сопоставляемых параметров. Однако в ней отсутствует возможность учета рассматриваемого элемента i со всеми $(M - 1)$ элементами дискретного ансамбля. Данный недостаток может быть устранен одним из способов, который рассмотрен в качестве теоретического в [4] и практического в [5].

При определении энтропии в (2) вводится симметричная неотрицательная вещественнозначная функция пар исходов $0 \leq \rho(x_i, x_j) = \rho_{ij} \leq 1$, которая будет

дополнять ансамбль (1). Данная функция на множестве вероятностного ансамбля позволяет определить «рандомизированное» расстояние ρ , подчиняющееся естественному условию $\rho_{ii} = 0$ и ограниченное сверху единицей.

Имея расстояние ρ , энтропия как априорная мера неопределенности дискретного вероятностного ансамбля с расстоянием определяется по формуле:

$$B_{\rho}(p) = -\sum_{i=1}^M p_i \log \sum_{j=1}^M (1 - \rho_{ij}) p_j, \quad (3)$$

где ρ_{ij} – расстояние между значениями событий i и j . В (3) энтропия получила название «геометризованная энтропия» или « B -энтропия».

Для определения указанных энтропий было разработано программное обеспечение на языке C#. Программа обладает следующими возможностями:

- выгрузка исходных данных из Excel;
- корректировка входных данных для оптимального соотношения прогнозируемых (плановых) и фактических показателей;
- построение диаграмм, демонстрирующих несоответствие показателей;
- построение гистограммы статистического распределения данных;
- выбор способа расчета и определение энтропии;
- вывод расчетных данных на экран и бумажный носитель.

Алгоритм определения энтропии поясним на примере. Энергосбытовая компания приобретает на оптовом рынке электроэнергию. Предварительно формирует план-прогноз на покупку электрической мощности на каждый час предстоящих суток. Для подготовки прогноза предприятие использует информационные технологии и экспертные оценки. По истечении суток фиксируется фактическое потребление электроэнергии. Данные о заявках на покупку и фактическом потреблении в течение апреля месяца представлены в табл. 1. С целью оценки соответствия прогнозных показателей с фактическими данными использовалось разработанное программное обеспечение [6].

Таблица 1 – Данные о планируемом и фактическом потреблении электроэнергии, МВтч

Дата	01.04.2011				...	30.04.2011			
Час суток	00	01	03	04	...	20	21	22	23
План	53,746	54,553	56,314	56,726	...	42,251	40,518	39,678	39,325
Факт	53,832	54,369	57,018	57,667	...	44,263	43,049	41,466	40,934

Всего рассматривалось 720 прогнозных и 720 фактических значений. На рис. 1 и 2 представлены, соответственно, прогнозные значения покупки электроэнергии на каждый час и фактического потребления, МВтч.

Определена разность отклонений прогнозных и фактических значений на каждый час, МВтч (рис. 3). По отклонениям, с учетом диапазона в размере 1 МВтч, построена гистограмма (рис. 4) и определены вероятности исходов событий.

Для каждого часа по формулам (2) и (3) получены расчетные значения энтропий (рис. 5), а также обобщенные величины энтропий за весь рассматриваемый период времени.

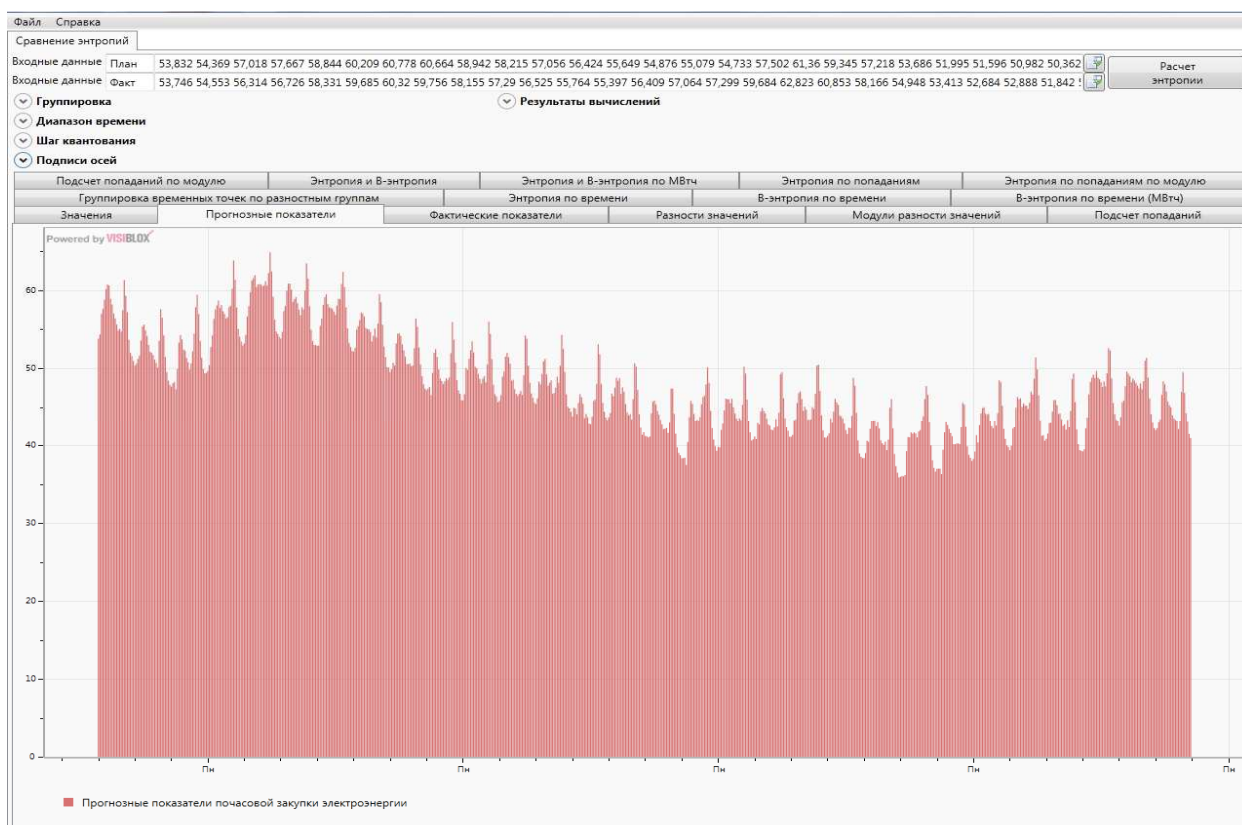


Рисунок 1. Диаграмма прогнозных показателей почасовой закупки электроэнергии, МВтч.

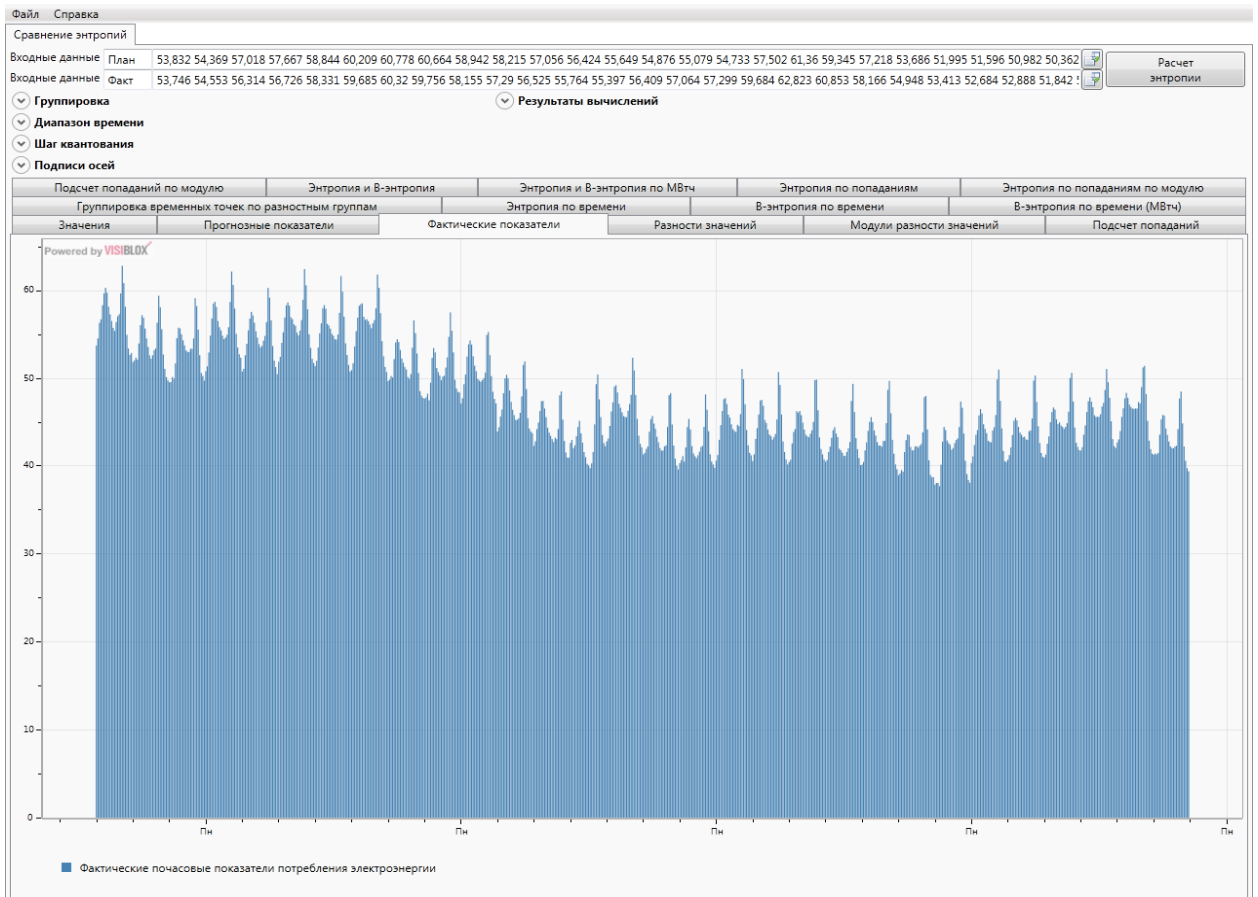


Рисунок 2. Диаграмма фактического почасового потребления электроэнергии.

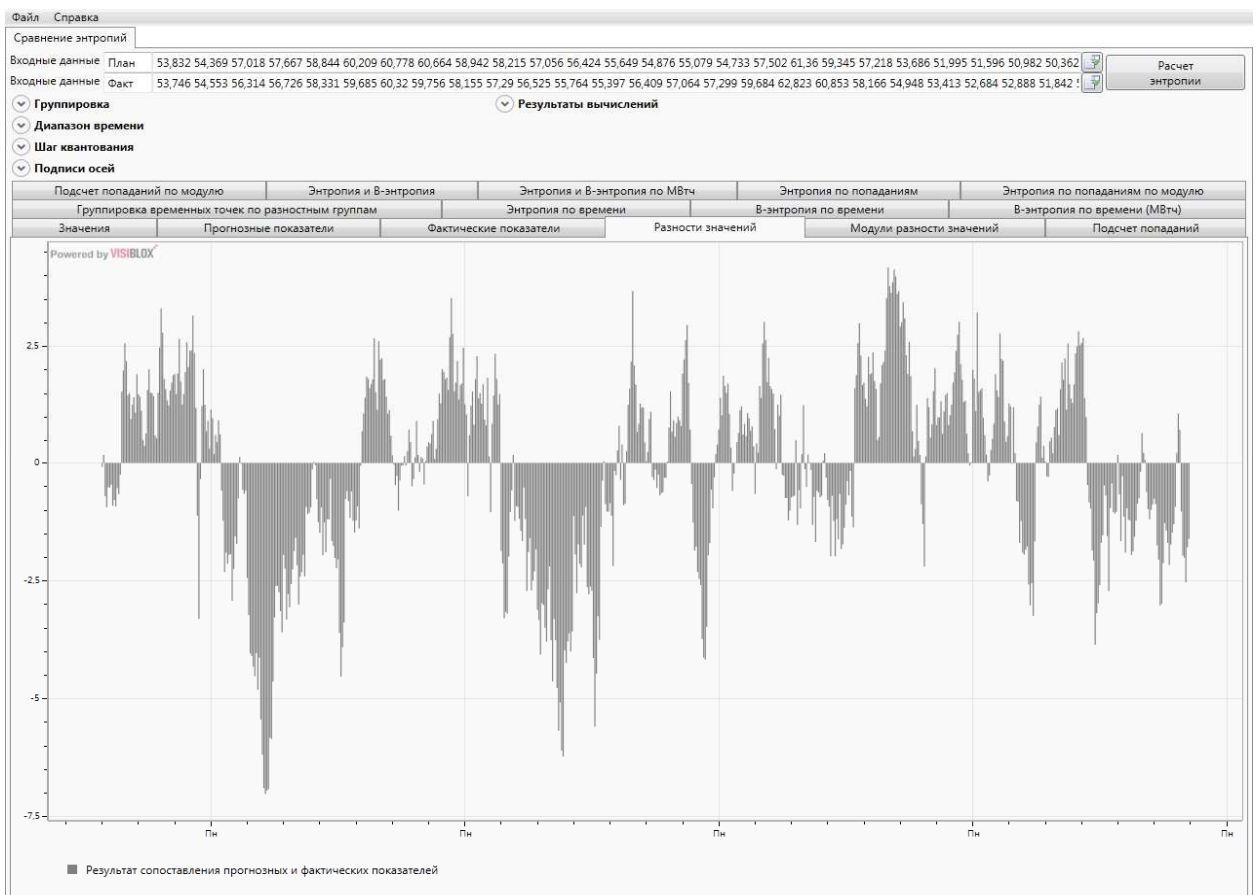


Рисунок 3. Диаграмма разностей прогнозных и фактических показателей за период времени.

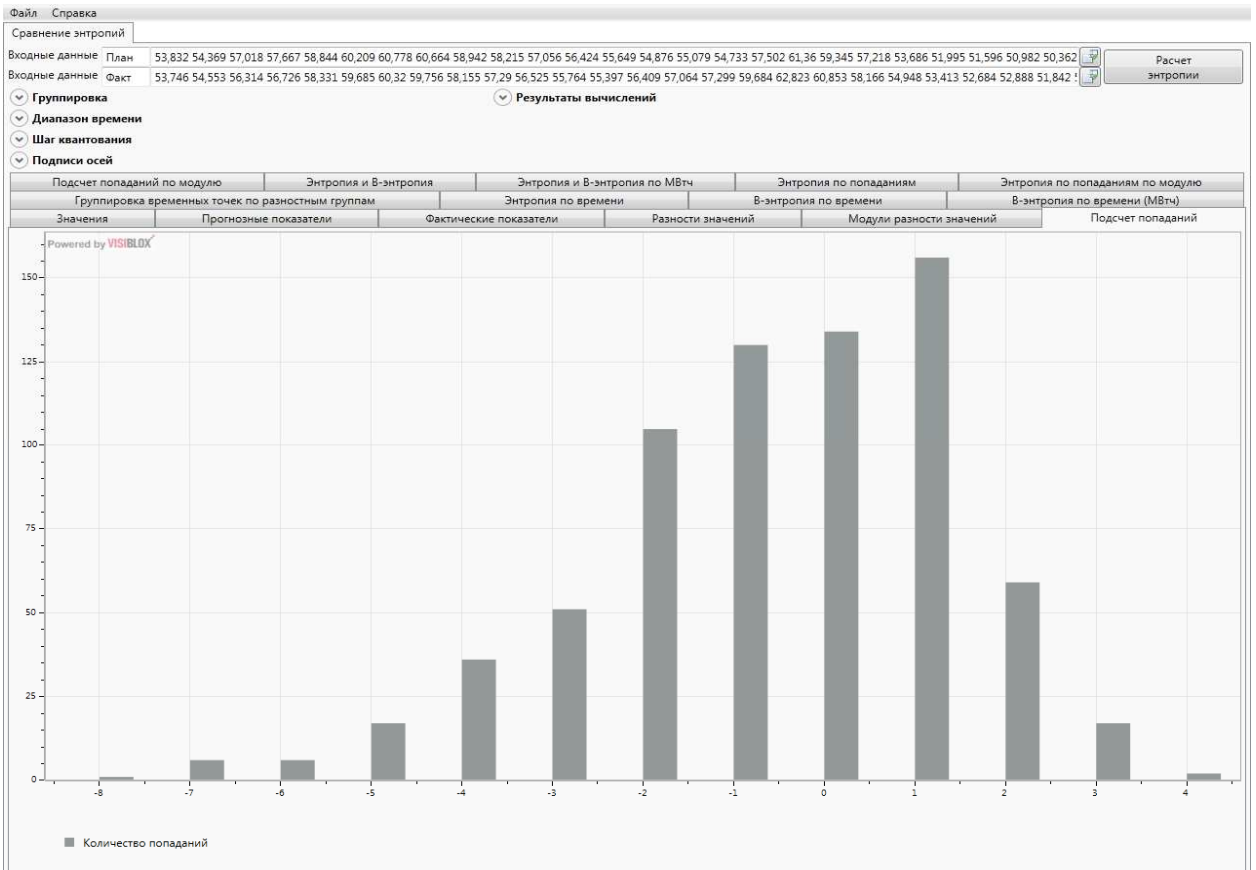


Рисунок 4. Гистограмма статистического распределения количества разностей по промежуткам.

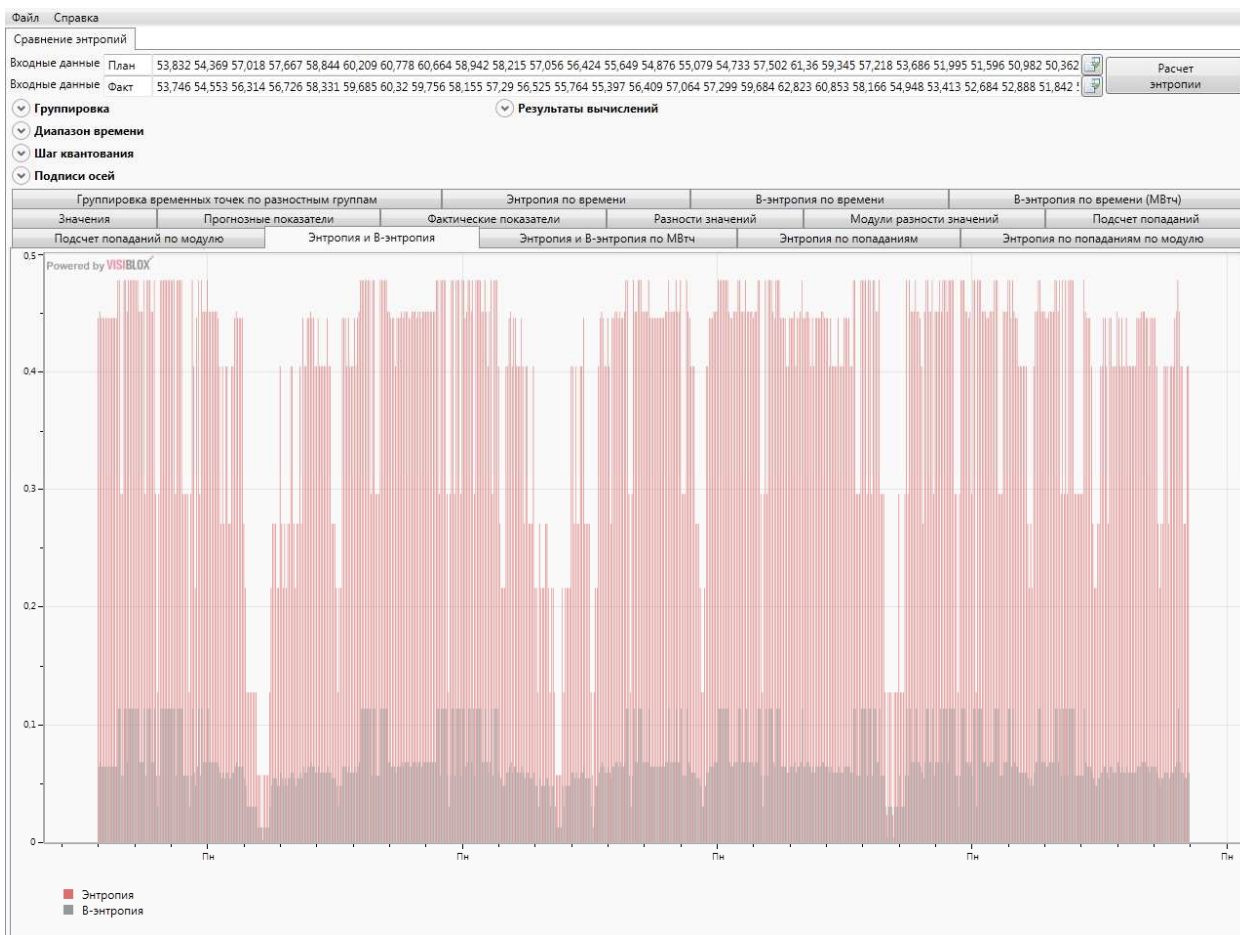


Рисунок 5. Диаграмма почасового сопоставления значений энтропий по формуле Шеннона и V-энтропии.

Динамика энтропий позволяет видеть «узкие места» при подготовке прогноза, сопоставлять между собой энтропии одного вида для интересующих нас моментов времени, а также энтропии разных видов для выбранного момента или промежутка времени.

Заключение. Произведя вычисления по формуле Шеннона (2) и V-энтропии (3), отметим наличие существенной разницы в полученных значениях. Соответственно с помощью V-энтропии можно с большей достоверностью учитывать не только исходы, но и соотношения (расстояния) между значениями для пар исходов.

Предложенный метод определения энтропии может быть полезен для сопоставления не только прогнозных, но и плановых показателей с фактическими данными.

С научной точки зрения V-энтропия привлекательна тем, что представляет собой перспективное направление энтропийного подхода в задаче оценки неопределенности данных. Дальнейшее изучение и развитие V-энтропийного подхода позволит вычислять неопределенность информации с большей точностью в любой прикладной задаче. Это позволит на практике рационально использовать ресурсы, время и другие составляющие.

При развитии теоретических исследований возможна разработка новых моделей на основе В-энтропийного подхода, которые окажут помощь в выработке эффективных решений по управлению бизнес-процессами.

Список литературы

1. Дулесов А.С., Ускова Е.А. Применение подходов Хартли и Шеннона к задачам определения количества информации технических систем // Вопросы современной науки и техники. Ун-т им. В.И. Вернадского. Сер. «Технические науки». – 2009. – № 2 (16).

2. Дулесов А.С., Ускова Е.А. Определение величины информации о структурном содержании технической системы из последовательно соединенных взаимозависимых элементов // Вопросы современной науки и техники. Ун-т им. В.И. Вернадского. Сер. «Технические науки». – 2009. – № 8 (22).

3. Дулесов А.С., Ускова Е.А. Применение статистической меры информации в задаче потребления энергии // Перспективы науки. Science prospects. – 2010. – № 1 [03].

4. Леус В.А. О геометрическом обобщении энтропии // Проблемы передачи информации. – 2003. – Т. 39. – Вып. 2. – С. 15–22.

5. Дулесов А.С., Швец С.В., Хрусталеv В.И. Применение формулы Шеннона и геометрического обобщения для определения энтропии // Перспективы науки. Science prospects. – 2010. – № 3 [05].

6. Коняев А.А., Хрусталеv В.И. Программа: вычисление энтропии ансамбля данных. Регистрационный номер в Фонде алгоритмов и программ PR11042. Дата регистрации в Фонде алгоритмов и программ: 2011-06-14.

Рецензенты:

Нагрузова Л.П., д.т.н., профессор кафедры «Строительство» Хакасского технического института – филиала ФГАОУ ВПО «Сибирский федеральный университет», г. Абакан.

Булакина Е.Н., д.т.н., профессор кафедры «Автомобили и автомобильное хозяйство» Хакасского технического института – филиала ФГАОУ ВПО «Сибирский федеральный университет», г. Абакан.
