

УДК: 519.8 (075.8)

**АНАЛИЗ КАЧЕСТВА ОЦЕНОК ПРОГНОЗОВ
С ИСПОЛЬЗОВАНИЕМ МЕТОДА КОМПЛЕКСИРОВАНИЯ**

П.И. БИДЮК, А.С. ГАСАНОВ, С.Е. ВАВИЛОВ

Используя оценки краткосрочных прогнозов, полученных различными методами, выполнено комплексирование оценок прогнозов на один шаг. При этом по вычисленным СКО для каждого метода (экспоненциальное сглаживание, четкий и нечеткий МГУА, фильтр Калмана) реализуется схема объединения оценок прогнозов, обеспечивающая увеличение качества прогнозирования в условиях правильного выбора весовых коэффициентов. Полученные экспериментальные результаты с применением вышеуказанных методов свидетельствуют о возможности повышения качества оценок прогнозов благодаря использованию предложенной схемы комбинирования в условиях, когда дисперсии ошибок оценок индивидуальных прогнозов близки между собой. Для сравнения результатов прогнозирования использовано множество статистических параметров качества прогноза.

ВВЕДЕНИЕ

Принятие высококачественных решений в экономике и финансах, а также при управлении техническими системами и технологическими процессами требует получения высококачественных прогнозов. Несмотря на наличие множества методов прогнозирования задача увеличения качества прогнозов остается актуальной, поскольку качество зависит от многих факторов, таких как качество и объем экспериментальных данных, наличие неопределенностей различного характера и типа исследуемых процессов (стационарные и нестационарные, линейные и нелинейные) [1–4]. Учесть все особенности процессов бывает очень трудно или невозможно из-за ряда объективных причин: малые выборки, большие шумы измерений, наличие существенных возмущений и больших выбросов. Поэтому необходимо искать методы и средства повышения качества моделей и прогнозов в условиях наличия упомянутых трудностей. В некоторой мере задача повышения качества прогнозов решается с помощью метода группового учета аргументов, нейронных сетей [3]. Однако использование этих методов имеет свои недостатки, обусловленные сложностью получаемых моделей и зачастую большими вычислительными затратами. В условиях наличия неопределенностей структурного, статистического и параметрического характера можно применять методы моделирования и прогнозирования на основе нечеткой логики и байесовских методов анализа данных. Как правило, использование этих

методов дает положительные результаты, однако их применение также сопряжено с определенными трудностями объективного и субъективного характера — это определение априорных распределений и выбор функций принадлежности.

Поэтому в данной работе ставится задача исследования возможности повышения качества оценок прогнозов путем объединения (комбинирования) оценок, полученных различными методами. Известно, что во многих случаях использование такого подхода дает возможность улучшить оценки кратко- и среднесрочного прогнозов.

Цель работы — управление качеством оценок прогнозов на базе вычисленного комбинированного прогноза с использованием разных методов прогнозирования — экспоненциального сглаживания, двойного экспоненциального сглаживания, полиномиального четкого и нечеткого МГУА и фильтра Калмана.

ПОСТАНОВКА ЗАДАЧИ

На основе месячных показателей о макроэкономических показателях Украины, а именно, процента изменения индекса потребительских цен (% ИПЦ) и национального внутреннего валового продукта (НВВП), а также данных о других макроэкономических показателях за период с 01.04.1996 по 01.07.1997 гг., построить прогнозирующие модели (ПМ) ЭС, двойного ЭС, моделей с использованием полиномиального МГУА на основе линейных и квадратичных описаний и фильтра Калмана. Провести экспериментальные исследования вышеперечисленных моделей и вычислить коэффициент комбинированного прогноза (анализ их точности, т.е. качества) для разных соотношений обучающей ($N_{обуч}$) и проверочной выборок ($N_{пров}$).

Пусть имеются методы (алгоритмы) прогнозирования A_1, A_2, \dots, A_n и пусть они используются для прогнозирования некоторого нелинейного случайного процесса $f(t)$.

Находим прогнозы по каждому из методов $y_{Ai}(t)$, $i = \overline{1, n}$, и пусть $\overline{S}_i^2 = \text{СКО}$ на проверочной выборке алгоритма A_i . Тогда комплексный прогноз по всем методам определяется согласно формуле [3]:

$$\hat{y}_k = \sum_{i=1}^n \frac{\overline{y}_{Ai}}{\overline{S}_i^2} / \sum_{i=1}^n \frac{1}{\overline{S}_i^2}. \quad (1)$$

РЕШЕНИЕ ЗАДАЧИ КОМБИНИРОВАНИЯ ОЦЕНОК ПРОГНОЗОВ, ПОЛУЧЕННЫХ РАЗНЫМИ МЕТОДАМИ

Усреднение прогнозов (весовые коэффициенты одинаковые)

Простейшим методом комбинирования оценок есть простое усреднение. Для двух методов прогнозирования среднее определяется просто:

$$\hat{y}_c(k) = \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2}, \quad (2)$$

где $\hat{y}_c(k)$ — комбинированный прогноз; $\hat{y}_1(k)$, $\hat{y}_2(k)$ — прогнозы, полученные разными методами. Если отдельные прогнозы не смещены (это должен обеспечивать метод прогнозирования), то комбинированный прогноз также будет несмещенным. Погрешность комбинированного прогноза определяется как среднее погрешностей отдельных оценок:

$$e_c(k) = y(k) - \hat{y}_c(k) = y(k) - \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2} = \frac{e_1(k) + e_2(k)}{2}, \quad (3)$$

где $y(k)$ — фактическое значение прогнозируемой переменной.

При этом дисперсия погрешности комбинированного прогноза определяется выражением:

$$\begin{aligned} \text{Var} \left[\frac{e_1(k) + e_2(k)}{2} \right] &= E \left[\frac{e_1(k) + e_2(k)}{2} \right]^2 = \frac{1}{4} E[e_1^2(k) + 2e_1(k)e_2(k) + e_2^2(k)] = \\ &= \frac{1}{4} \{ E[e_1^2(k)] + 2E[e_1(k)e_2(k)] + E[e_2^2(k)] \} = \\ &= \frac{1}{4} \left[\sigma_1^2 + 2 \frac{E[e_1(k)e_2(k)]}{\sigma_1 \sigma_2} \sigma_1 \sigma_2 + \sigma_2^2 \right] = \frac{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}{4}. \end{aligned}$$

Таким образом, окончательно дисперсия комбинированного прогноза определяется по формуле:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4}, \quad (4)$$

где ρ — коэффициент корреляции между погрешностями прогнозов. Если погрешности оценок прогнозов, полученных по двум моделям, независимы, то формула (4) упрощается:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2}{4}. \quad (5)$$

Таким образом, если дисперсии близки по значениям и погрешности оценок прогнозов независимы, то дисперсия комбинированной погрешности будет значительно меньше любой из двух дисперсий. Например, пусть

$$\sigma_c^2 = \frac{100 + 100}{4} = 50.$$

Но даже при существовании достаточно высокой корреляции между погрешностями ошибок прогнозирования дисперсия погрешности комбинированного прогноза будет меньшей, чем дисперсия каждого метода отдельно. Например, пусть $\sigma_1^2 = \sigma_2^2 = 100$ и $\rho = 0,8$. Тогда получаем

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{100 + 100 + 2 \cdot 0,8 \cdot 10 \cdot 10}{4} = 90.$$

Даже в этой ситуации наблюдается уменьшение дисперсии погрешности прогноза после усреднения оценок, полученных по двум методам.

Однако ситуация изменяется в случае, когда дисперсии индивидуальных погрешностей значительно отличаются. Например, пусть $\sigma_1^2 = 100$, $\sigma_2^2 = 16$ и $\rho = 0,8$, тогда:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{100 + 16 + 2 \cdot 0,8 \cdot 10 \cdot 4}{4} = 45.$$

Таким образом, если дисперсии погрешностей значительно отличаются одна от другой, то простое усреднение результатов делать не нужно. Теперь можно сделать вывод: простое усреднение можно применять с целью улучшения качества прогнозов в случаях, когда дисперсии индивидуальных погрешностей прогнозирования приблизительно равны или не очень отличаются по своим значениям.

Взвешенное комбинирование прогнозов

Если информация относительно характеристик индивидуальных оценок прогнозов отсутствует, то можно присвоить разные весовые коэффициенты отдельным прогнозам на основе субъективных или экспертных суждений:

$$\hat{y}_c(k) = w_1 \hat{y}_1(k) + w_2 \hat{y}_2(k),$$

где w_1, w_2 — весовые коэффициенты. Очевидно, что большие значения весовых коэффициентов необходимо присваивать тем индивидуальным прогнозам, которые имеют меньшую дисперсию погрешностей. При этом для корректности вычислений необходимо, чтобы выполнялось условие $w_1 + w_2 = 1$.

Как правило, погрешности прогнозов для конкретных моделей и процессов известны или их можно определить на обучающей выборке. Это дает возможность объективно подойти к решению задачи выбора весовых коэффициентов. Поскольку модели, которые дают меньшие суммы квадратов погрешностей прогнозов, генерируют более качественные прогнозы, то логично принять эту меру за основу для определения весовых коэффициентов. Обозначим сумму квадратов погрешностей прогнозирования (для исторического прогноза) через

$$sse = \sum_{k=1}^N e^2(k).$$

Теперь можно записать выражения для весовых коэффициентов отдельных прогнозов [5]:

$$w_1 = \frac{1/sse_1}{1/sse_1 + 1/sse_2},$$

$$w_2 = \frac{1/sse_2}{1/sse_1 + 1/sse_2},$$

где sse_1, sse_2 — суммы квадратов погрешностей для каждого из используемых методов.

Например, пусть $sse_1 = 100$, $sse_2 = 16$:

$$w_1 = \frac{1/100}{1/100 + 1/16} = \frac{0,01}{0,01 + 0,0625} = 0,1379,$$

$$w_2 = \frac{1/16}{1/100 + 1/16} = \frac{0,0625}{0,01 + 0,0625} = 0,8621.$$

То есть, объективно присвоен больший весовой коэффициент более точной оценке прогноза. При этом выполняется условие $\sum w_i = 1$, которое необходимо для достижения корректности применения метода [6, 7].

Таким образом, можно сделать общий вывод, что при прогнозировании процессов произвольной природы необходимо применять как отдельные методы, так и комбинирование оценок прогнозов, вычисленных с помощью разных методов. При этом весовые коэффициенты для отдельных оценок можно получить разными способами, что также способствует поиску лучшего варианта формирования прогнозов. Очевидно, что такие подходы к прогнозированию лучше реализовывать в соответствующих системах поддержки принятия решений с автоматизацией функций обработки данных, оценивания структур и параметров моделей и прогнозов на их основе.

В соответствии с предложенным методом комбинирования получим схему вычислений в следующем виде, представленном на рис. 1, а на рис. 2 — схему деления данных на обучающую и тестовую выборки.

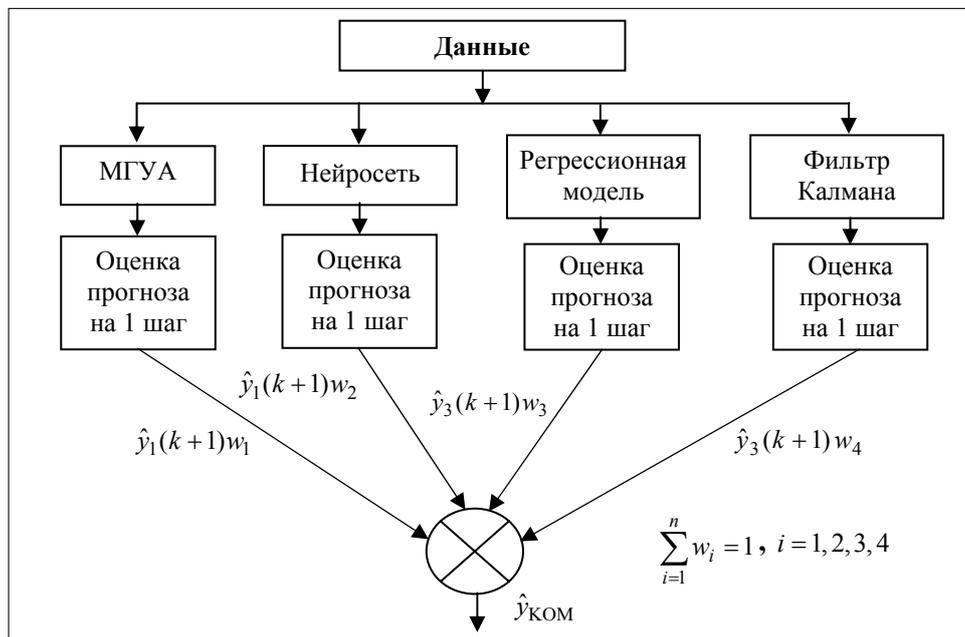


Рис. 1. Схема получения комбинированного прогноза

Комбинированный прогноз для четырех методов вычисляется по формуле:

$$\hat{y}_{КОМ}(k+1) = \sum_{i=1}^4 w_i \hat{y}_i(k+1).$$

СКО — среднеквадратическая ошибка $\sum_1^{100} e_i^2(k)$ — на обучающей выборке из ста значений используется для вычисления весовых коэффициентов следующим образом:

$$w_i = \frac{\sum_1^{100} e_i^2(k)}{\sum_1^{100} e_1^2(k) + \sum_1^{100} e_2^2(k) + \dots + \sum_1^{100} e_4^2(k)}$$

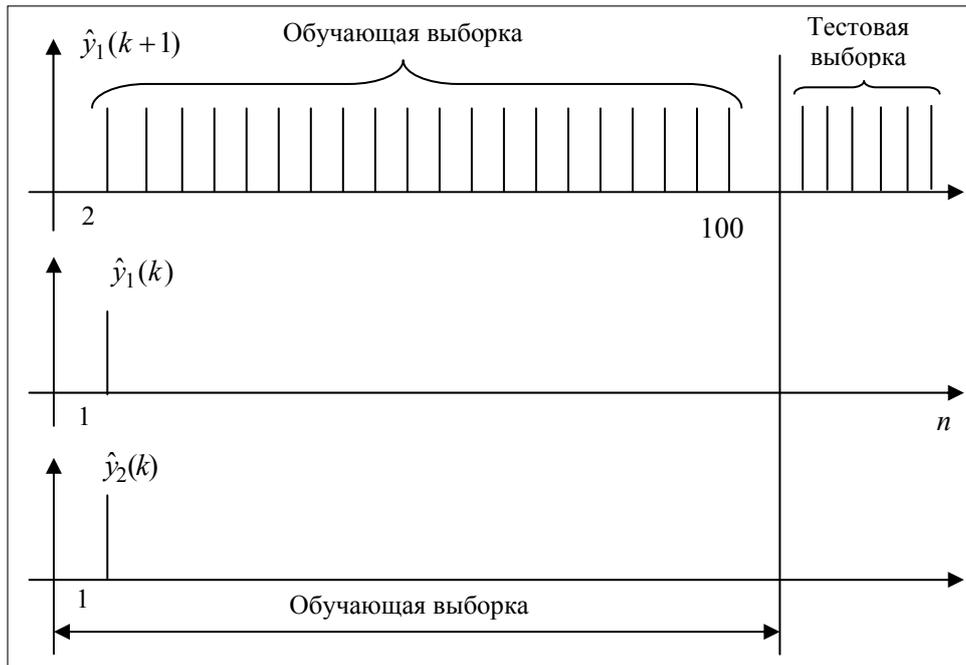


Рис. 2. Деление исходных данных на обучающую и тестовую выборки

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

При выполнении экспериментальных исследований варьировались соотношения между объемом обучающей и проверочной выборок.

В табл. 1 представлен фрагмент вычисленных значений статистических характеристик качества оценок прогнозов, вычисленных разными методами прогнозирования, а именно: СКО (RSME), средняя ошибка (ME), средняя ошибка в процентах (MPE) и средняя абсолютная ошибка в процентах (MAPE).

На рис. 3 приведены графики прогнозов, полученные разными методами прогнозирования.

В табл. 2 приведены значения прогнозов и СКО на один шаг, полученные с помощью различных методов прогнозирования для иллюстрации качества прогнозирования без комплексирования.

На рис. 4 представлен внешний вид интерфейсного окна реализованной программы.

В табл. 3 представлен фрагмент вычисленных значений прогноза на один шаг, выполненных разными методами прогнозирования (а также СКО (RSME), ME, MPE, MAPE) для другого ряда данных (валовый внутренний продукт) при делении выборки на обучающую и проверочную в таком соотношении $N_{\text{обуч}} = 65$ и $N_{\text{пров}} = 35$.

Таблица 1. Характеристики качества оценок прогнозов для выбранных методов

Критерий	У эксп	У эксп2	У МГУА	У НМГУА	У Калман
RSME	0,215812	0,379131	0,115582	0,147774	0,400863
ME	0,061722	0,09459	-0,01399	0,055247	0,302063
MPE	0,007206	0,011143	-0,00182	0,006221	0,035074
MAPE	0,019299	0,036307	0,008985	0,014457	0,039612
U	0,013533	0,023818	0,007213	0,009265	0,025551
U_M	10,25129	6,934964	27,36097	11,1058	4,919036

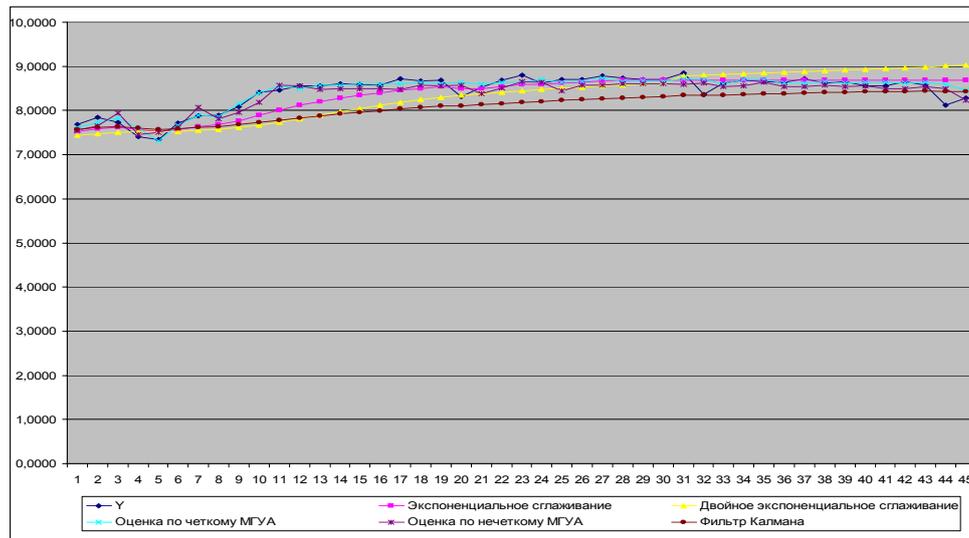


Рис. 3. Графики прогнозов, полученные разными методами прогнозирования

Таблица 2. Значения прогнозов и СКО на один шаг, полученные разными методами прогнозирования

Данные	Методы прогнозирования				
	У эксп	У эксп2	У МГУА	У НМГУА	У Калман
y_i	$\hat{y}_1(k+1)$	$\hat{y}_2(k+1)$	$\hat{y}_3(k+1)$	$\hat{y}_4(k+1)$	$\hat{y}_5(k+1)$
7,6780	7,5222	7,4413	7,6184	7,5555	7,5651
RSME	0,215812	0,379131	0,115582	0,147774	0,400863

На рис. 5 приведены графики прогнозов, полученные разными методами прогнозирования.

Результаты обработки интегрированной системой прогнозирования (комбинирование результатов разных методов с весами, обратно пропорциональными дисперсии).

Рис. 4. Пример внешнего вида интерфейсного окна

Таблица 3. Значения прогнозов и СКО на один шаг, полученные разными методами прогнозирования для ряда финансовых данных

Y	Y эксп	Y эксп2	Y МГУА	Y НМГУА	Y Калман
RSME	0,015417	0,007754	0,008807	0,006197	0,012185
ME	-0,00929	-0,00133	0,004194	0,002313	-0,00704
MPE	-0,00768	-0,00113	0,003451	0,001885	-0,00582
MAPE	0,009571	0,005175	0,005185	0,004046	0,007952
U	0,006507	0,003284	0,003739	0,002629	0,005148
U_M	5,330868	7,143253	8,6459	8,476586	5,22148

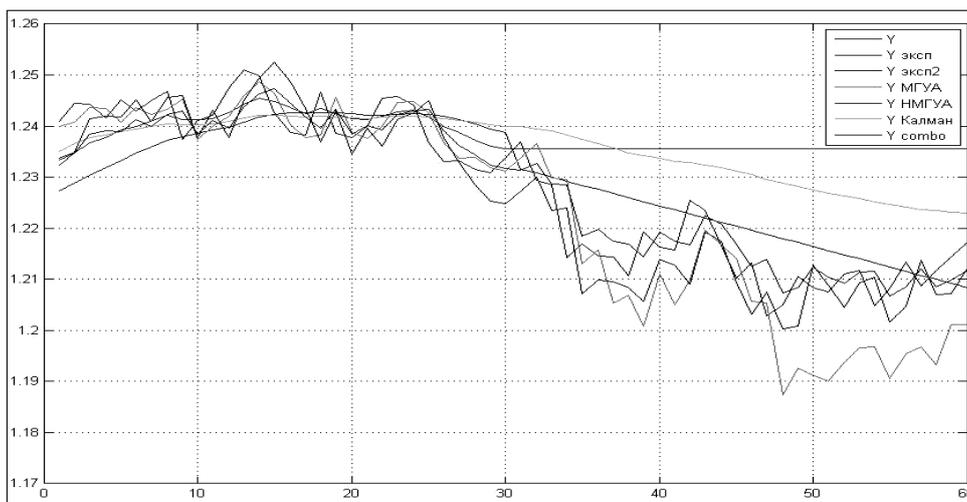


Рис. 5. Графики прогнозов, полученные различными методами прогнозирования

Веса для отдельных прогнозов:

Y эксп	Y эксп2	Y МГУА	Y НМГУА	Y Калман
0,0636182	0,249936	0,194041	0,390993	0,101413

Значения основной переменной (Y) и их оценки представлены в табл. 4. Фрагмент вычисленных значений прогноза на один шаг, выполненных разными методами прогнозирования, и значения комбинированного прогноза.

СКП (RSME):

Y эксп	Y эксп2	Y МГУА	Y НМГУА	Y Калман	Y combo
0,0160133	0,00807901	0,00916907	0,00645933	0,0126831	0,00507773

Таблица 4. Значения прогнозов и СКО на один шаг, полученные различными методами прогнозирования

Y	Y эксп	Y эксп2	Y МГУА	Y НМГУА	Y Калман	Y combo
1,241	1,2324	1,2274	1,2399	1,2338	1,235	1,23342
1,2445	1,2348	1,2289	1,2407	1,2349	1,2366	1,23469
1,2443	1,2367	1,2304	1,2437	1,2415	1,2377	1,23846
1,2416	1,2377	1,2319	1,2434	1,2419	1,2382	1,23905
1,2452	1,2392	1,2333	1,2408	1,2417	1,239	1,23899
1,243	1,2399	1,2347	1,2437	1,2452	1,2394	1,24136
1,2449	1,2409	1,2359	1,2423	1,241	1,2399	1,23986
1,2106	1,2355	1,2172	1,1926	1,2008	1,2282	1,20829
1,2083	1,2355	1,2164	1,1912	1,2127	1,2275	1,2124
1,2075	1,2355	1,2156	1,19	1,2089	1,2269	1,21042
1,2111	1,2355	1,2148	1,1937	1,2045	1,2263	1,20916
1,2117	1,2355	1,214	1,1966	1,2093	1,2259	1,21136
1,2048	1,2355	1,2132	1,1967	1,2104	1,2252	1,21154
1,2082	1,2355	1,2124	1,1906	1,2017	1,2246	1,20669
1,2134	1,2355	1,2116	1,1955	1,2046	1,2242	1,20854
1,2087	1,2355	1,2108	1,1967	1,2138	1,2237	1,21212
1,2117	1,2355	1,21	1,1933	1,2071	1,2234	1,20861
1,2144	1,2355	1,2092	1,2012	1,2072	1,2231	1,20995
1,2172	1,2355	1,2084	1,2011	1,2122	1,2229	1,21166

Теперь убираем один из методов прогнозирования, например, на основе фильтра Калмана. В результате получены такие весовые коэффициенты для индивидуальных прогнозов:

Y эксп	Y эксп2	Y МГУА	Y НМГУА
0,0707981	0,278143	0,21594	0,435119

Значения СКО для отдельных методов и комбинированного прогноза:

Y эксп	Y эксп2	Y МГУА	Y НМГУА	Y combo
0,0160133	0,00807901	0,00916907	0,00645933	0,00507872

При комбинировании трех методов получены следующие весовые коэффициенты:

$Y_{\text{эксп}}$	$Y_{\text{эксп}2}$	$Y_{\text{МГУА}}$
0,125333	0,492392	0,382275

Среднеквадратическая ошибка для этого случая комбинирования оценок прогнозов:

СКП (RSME):

$Y_{\text{эксп}}$	$Y_{\text{эксп}2}$	$Y_{\text{МГУА}}$	Y_{combo}
0,0160133	0,00807901	0,00916907	0,00567565

Таким образом, качество оценок комбинированных прогнозов уменьшается с уменьшением количества используемых методов.

ВЫВОДЫ

Проведенные экспериментальные исследования позволяют сделать следующие выводы: с увеличением методов прогнозирования качество комбинированного прогноза улучшается и стремится к точному значению; по прогнозным характеристикам для всех соотношений обучающей и проверочной выборок минимальные значения по показателям СКО и САПП полностью зависят от качества отдельных прогнозов, полученных индивидуальными методами; во всех экспериментах модель четкого полиномиального МГУА с линейными частичными описаниями занимает второе место.

Таким образом, проведенный анализ показал, что качество прогноза улучшается с увеличением количества методов прогнозирования. Прогнозирование выполнено для макроэкономических показателей (ИПЦ и НВВП) экономики Украины. Рассмотренный подход можно использовать для решения задачи повышения качества прогнозов в системах автоматического и автоматизированного управления, в эргатических системах автоматизированного контроля технологических параметров и т.д.

В дальнейших исследованиях целесообразно рассмотреть возможности комбинирования идеологически разных методов — регрессионный анализ и нечеткую логику, нейронные сети и байесовский анализ данных. А также целесообразно построить систему поддержки принятия решений на основе альтернативных методов.

ЛИТЕРАТУРА

1. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учебное пособие. — М.: Финансы и статистика, 2002. — 411 с.
2. Бидюк П.И., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів. — Київ: Політехніка, 2012. — 360 с.
3. Зайченко Ю.П. Основы проектирования интеллектуальных систем. — К.: Видавничий дім «Слово», 2004. — 352 с.
4. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление (выпуск 1). — М.: Мир, 1974. — 408 с.
5. Tsay R.S. Financial time series analysis. — Chicago: John Wiley & Sons, Inc., 2002. — 455 p.
6. Poon S.H. A Practical guide to forecasting financial time volatility. — NY: John Wiley & Sons, Inc., 2005. — 238 p.
7. Грешилов А.А., Стакун В.А., Стакун А.А. Математические методы построения прогнозов. — М.: Радио и связь, 1997. — 112 с.

Поступила 27.02.2013