

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ СРЕДСТВАМИ MS SQL SERVER

Шаркова М.А.

научный руководитель Сарапулова Т.В.

Кузбасский государственный технический университет имени Т.Ф. Горбачева

В настоящее время информация является как основным «оружием», так и основным «орудием труда». Естественно появляется необходимость в средствах, способных анализировать, группировать и обрабатывать огромное количество разрозненных данных. Для решения подобных задач применяется интеллектуальный анализ данных (англ. Data Mining), целью которого является обнаружение неявных закономерностей в наборах данных. Одним из возможных вариантов проведения такого анализа является использование средств интеллектуального анализа данных в СУБД Microsoft SQL Server службами Analysis Services.

Службы Analysis Services предоставляют следующие функции и средства для создания решений по интеллектуальному анализу данных:

- набор стандартных алгоритмов интеллектуального анализа данных;
- конструктор интеллектуального анализа данных, предназначенный для создания и просмотра моделей интеллектуального анализа данных, управления ими и построения прогнозов;
- язык расширений интеллектуального анализа данных (Data Mining Extensionsto SQL, DMX).

Для работы с предоставляемыми средствами интеллектуального анализа используется среда Business Intelligence Development Studio. Также SQL Server поддерживают создание, управление и использование моделей интеллектуального анализа данных из Microsoft Excel с помощью Надстроек интеллектуального анализа данных SQL Server для MS Office.

Одна из наиболее востребованных задач интеллектуального анализа данных – задача прогнозирования. В условиях нестабильного рынка, прогноз необходим как на финансовых и фондовых рынках, так и на предприятиях, в органах государственной власти, в научных организациях и т.д.

Для решения данной задачи используется инструмент Прогноз (англ. Forecast), позволяющий построить прогноз значений числового ряда. При этом учитываются имеющиеся тенденции (тренды), сезонность, другие факторы. Классическим примером является прогнозирование цен акций на бирже или курса денежных единиц.

Источником данных для прогноза может служить электронная таблица Excel. Для примера возьмем таблицу с двумя числовыми последовательностями, возможно связанными между собой (рис. 1).

	A	B	C	D	E
4	Пример прогнозирования				
5					
6		Дата	Курс	Курс 2	
7		10.май	30,2306	35,7514	
8		11.май	30,1793	34,9825	
9		12.май	30,2652	35,0785	
10		13.май	30,3299	35,2457	
11		14.май	30,9758	35,4576	
12		15.май	30,9417	34,2754	
13		16.май	31,3921	35,3333	
14		17.май	31,1582	34,8656	
15		18.май	31,0644	35,4242	
16		19.май	31,3803	35,1675	
17					
18					

Рис. 1. Образец данных для прогнозирования

Инструмент Прогноз ищет в анализируемой таблице три типа шаблона:

1. Тренд – тенденцию изменения значений
2. Периодичность (сезонность) – событие повторяется через определённые интервалы
3. Взаимная корреляция – зависимость значений одного ряда от других

Настройка параметров начинается с выбора анализируемых столбцов и количества прогнозируемых единиц времени. Анализируемые столбцы должны содержать числовые или денежные значения. Следующим шагом должен быть выбор временной отметки. Значения этого столбца должны быть уникальны. В нашем случае, очевидно, что это столбец «Дата» (рис. 2).

Что касается периодичности, то предлагаются следующие варианты: определить автоматически, ежечасно, ежедневно, еженедельно, ежемесячно, ежеквартально, ежегодно. Если периодичность не известна, то рекомендуется выбирать «определить автоматически», чтобы инструмент проверил последовательность на наличие периодичности разных типов.

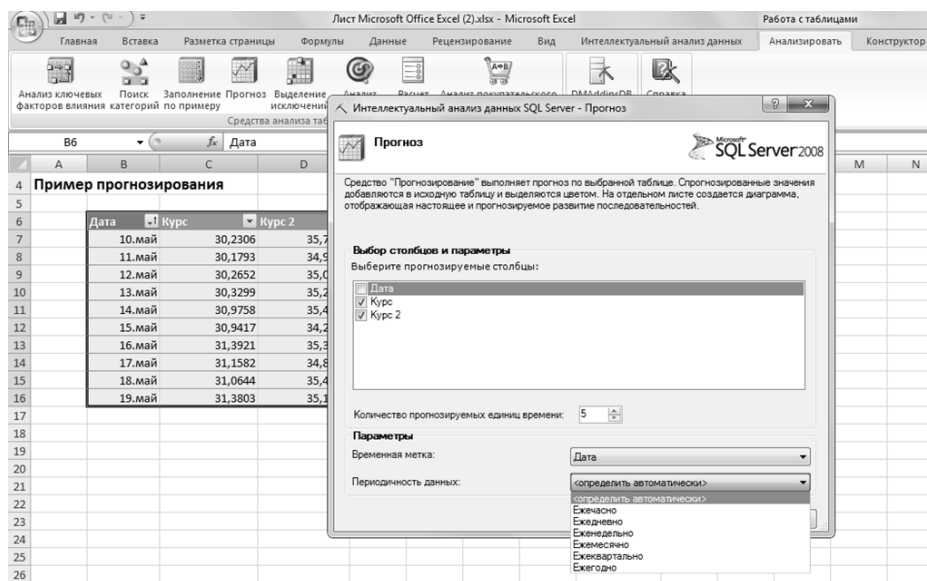


Рис. 2. Настройка параметров

По завершению настройки параметров создается отчет с графиком, где непрерывной линией изображаются исходные значения, а пунктирной – спрогнозированные. В исходную таблицу добавляются результаты прогноза (рис. 3).

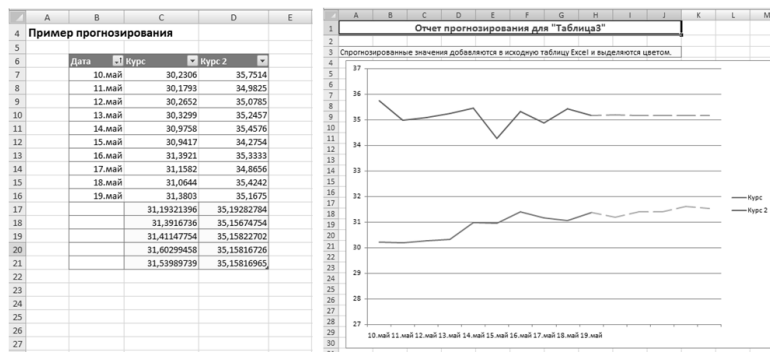


Рис. 3. Результат прогноза

Прогнозирование не дает 100% точности, однако при достаточном количестве данных является доступным и быстрым способом получить необходимые данные для дальнейшего планирования деятельности.