

Использование статистики eFind.ru для прогнозирования спроса на электронные компоненты

Алексей СЛАВГОРОДСКИЙ,
к. ф.-м. н.
alexey@efind.ru

Мы продолжаем цикл статей, посвященных анализу российского рынка электронных компонентов (ЭК) на основе статистики поисковой системы eFind.ru. Поисковые запросы к eFind.ru представляют собой наименования ЭК, а результатом поиска является информация о наличии и стоимости данного компонента на российском рынке. С помощью анализа этих запросов можно исследовать спрос на ЭК на российском рынке. В наших предыдущих статьях [1–3] статистика поисков использовалась для анализа уже произошедших изменений спроса на компоненты. В этой статье мы предприняли попытку сформулировать некоторые «симптомы» в поведении статистики поисков, появление которых могло бы свидетельствовать о росте спроса на тот или иной компонент.

Признаки роста

Мы взяли статистику поисков за 2008–2013 годы и для каждого запроса проследили изменения ежемесячного числа соответствующих поисков. Для снижения роли случайных флуктуаций данные были сглажены по методу скользящего среднего. Затем были выделены те запросы, для которых максимальное число поисков в месяц превышало 10 (здесь и далее речь идет о сглаженном числе поисков, если иное не указано особо). Всего было выделено 3890 запросов.

У каждого из этих запросов мы искали период роста числа поисков и фиксировали интервал времени, в течение которого ежемесячное число поисков удваивалось. Выяснилось, что более чем в 60% случаев для удваивания числа поисков требуется не более девяти месяцев. Именно этот срок и был выбран в качестве наиболее характерного времени роста.

Как определить, что, с некоторой вероятностью, начавшийся рост числа поисков, посвященных какому-то компоненту, в течение ближайших месяцев приведет к удваиванию этого числа? Была поставлена задача определить признаки, которые позволят сделать такое предположение. Имеющиеся у нас статистические данные за прошедшие годы позволяли проверить наши гипотезы и выбрать те, которые, если бы в свое время были применены, показали бы лучшие результаты по точности прогноза.

Задача была сформулирована следующим образом: мы видим, что на протяжении M месяцев число поисков по определенному запросу

растет не менее чем на $N\%$ в месяц. Какова вероятность того, что мы находимся в начале девятимесячного интервала, в течение которого число поисков удвоится? В таблице 1 представлены результаты обработки имеющихся у нас статистических данных. Указана доля случаев, когда наша гипотеза оказывалась верна, а также число запросов, чей рост удалось бы с ее помощью предсказать.

Видно, что максимальная точность прогноза (98%) достигается для случая I , когда мы на протяжении трех месяцев наблюдаем рост не менее чем на 20% каждый месяц. Это вполне логично, так как за эти три месяца число поисков уже увеличилось не менее чем на 70%, то есть до желаемых 100% роста либо осталось немного, либо этот рост уже состоялся. С другой стороны, далеко не всегда число поисков растет так быстро. В данном случае мы бы смогли спрогнозировать рост только для 369 запросов (из примерно 2000, для которых такой рост в итоге состоялся).

Максимальное число предсказанных запросов (1809) мы имели в случае A , если бы отслеживали ситуации, когда месячное число поисков по какому-то запросу окажется не менее чем на 10% больше, чем число поисков по этому запросу за предыдущий месяц. Однако точность таких прогнозов составила бы 20%. То есть в четырех случаях из пяти рост числа поисков в каком-то одном месяце не означал бы начала «большого» роста.

Некоторый баланс между точностью прогноза и типичностью признаков, на основе которых делается прогноз, наблюдается в случаях B и E . Точность прогноза составляет 41% (предсказан двукратный рост для 1435 запросов из примерно 2000) и 44% (1525) соответственно.

Формально же наиболее удачным «симптомом» начинающегося удваивания числа поисков можно назвать случай H , когда на протяжении трех месяцев наблюдается рост не менее чем на 10% ежемесячно, так как в этом случае произведение точности прогноза на число предсказанных запросов максимально.

Некоторые примеры из статистики прошлых лет

В одной из наших предыдущих статей [3] анализировалась статистика поисков по компонентам, которые только выходят на рынок. Эти компоненты определенно должны были в какой-то момент

Таблица 1. Оценка точности прогнозов, основанных на различных признаках роста

Случай	Признаки роста, %	Точность прогноза, %	Число точных предсказаний
A	Число поисков по сравнению с предыдущим месяцем увеличилось не менее чем на	10	1809
B		20	1435
C		30	1010
D		40	642
E		10	1525
F	Число поисков на протяжении двух месяцев подряд ежемесячно растет не менее чем на	20	855
G		30	353
H	Число поисков на протяжении трех месяцев подряд ежемесячно растет не менее чем на	10	1148
I		20	369

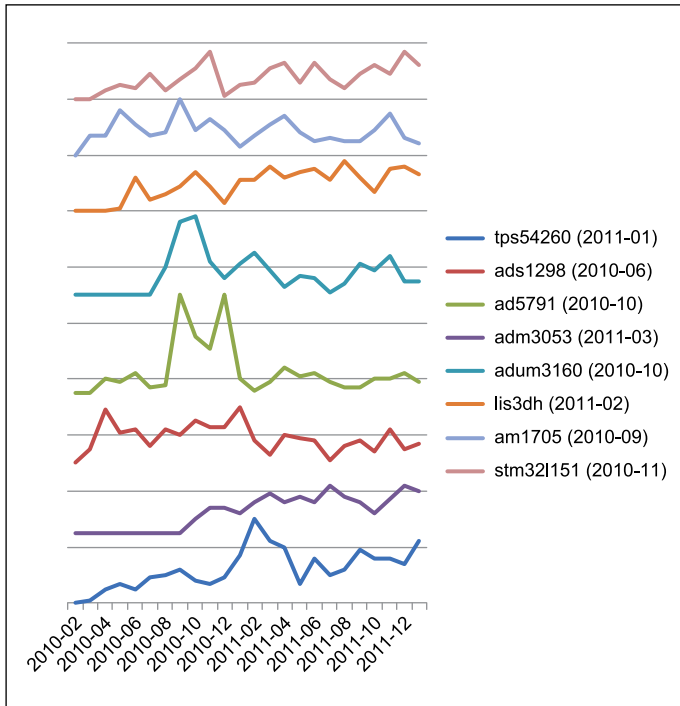


Рисунок. Динамика числа поисков по компонентам, для которых прогнозировался рост

иметь положительную динамику числа поисков, так как изначально это число равнялось нулю. Мы применили описанные выше методы прогнозирования к компонентам из [3]. На рисунке показано, как менялось число поисков (не сглаженное!) по этим запросам. В скобках рядом с наименованием компонента указан месяц, в который для этого компонента был бы предсказан дальнейший рост.

Детали расчета:

- Не учитывались месяцы, когда число поисков было меньше пяти (чтобы, например, игнорировать «многократный рост» числа поисков от одного до четырех в месяц).
- Использовались методы *B*, *E*, *H*.
- Указан самый ранний момент прогноза. (В некоторых случаях рост числа поисков по данному запросу предсказывался всеми тремя методами, но по методу *B* сигнал о росте поступал раньше.)

Для tps54260 и adm3053 действительно наблюдается рост, который продолжается как минимум год после момента формулирования прогноза. Для adum3160, ad5791 и ads1298 также есть периоды с повышенным уровнем запрашиваемости, но длительность этих периодов составляет всего несколько месяцев, после чего рост сменяется спадом. Момент прогноза для adum3160 вообще совпал с максимумом, то есть для этого компонента гипотеза о росте числа поисков была выдвинута в тот момент, когда реально число поисков в дальнейшем будет только падать. (Хотя формально сам факт удваивания числа поисков зафиксирован.) Для остальных запросов (lis3dh, am1705, stm321151) никакой ярко выраженной динамики не наблюдается. Запрашиваемость этих компонентов достигла определенного уровня, вблизи которого и колеблется.

Эти результаты вполне укладываются в оценки из таблицы 1, в которых точность используемых методов прогноза оценивается как 40–70%.

Текущий прогноз

В настоящий момент мы оперируем статистикой поисков с января 2008-го по июнь 2013 года включительно. В таблице 2 приводятся запросы, для которых в июне 2013 года был предсказан рост числа поисков. В ближайшем будущем можно будет проверить правильность этого прогноза.

Таблица 2. Запросы, для которых в июне 2013 года был предсказан дальнейший рост числа поисков

Метод В	Метод Е	Метод Н
irf3205	spw47n60c3	mnp30036
smbj50a	bd7820	
sc552904csp	avrpx128a1	
tip122	stm32f429	
uc3843	lsm303dlhc	
hmc6343		

Заключение

Мы сформулировали несколько признаков того, что число поисков по определенному электронному компоненту устойчиво растет вплоть до удваивания этого количества по сравнению с текущими показателями. По нашим оценкам, точность такого прогноза составляет 40–70%. Мы считаем, что приведенная в статье методика представляет интерес для поставщиков электронных компонентов, поскольку она может помочь в планировании складских запасов.

Литература

1. Славгородский А. Анализ спроса на продукцию отечественных производителей электронных компонентов // Компоненты и технологии. 2012. № 10.
2. Славгородский А. Анализ спроса на электронные компоненты в 2012 году // Компоненты и технологии. 2013. № 4.
3. Славгородский А. Об интересе аудитории eFind.ru к новинкам рынка электронных компонентов // Компоненты и технологии. 2011. № 10.