

ПРОГНОЗИРОВАНИЕ ВОСТРЕБОВАННОСТИ СПЕЦИАЛЬНОСТИ С ИСПОЛЬЗОВАНИЕМ ГИБРИДНОЙ ЭКСПЕРТНОЙ СИСТЕМЫ

М.В. Томашев

В настоящее время вследствие демографического кризиса перед ВУЗаами все острее встает вопрос о наборе необходимого количества абитуриентов. Серьезные исследования в этой области многие ВУЗы не проводят, полагаясь на опыт прошлых лет и на интуицию руководителей. Однако такой подход для современного рынка не приемлем, необходимы серьезные маркетинговые исследования с применением информационных технологий.

Учитывая присутствие ВУЗа одновременно на образовательном рынке и рынке труда, основными направлениями прогнозирования, являются востребованность специальности со стороны абитуриентов и востребованность специальности на рынке труда. Определение последней является особенно сложным, ведь необходимо прогнозирование на 4-5 лет вперед, т.к. набирая абитуриентов сегодня, должен гарантироваться большой процент трудоустройства по специальности в будущем. Выполнение данной задачи крайне трудно осуществить, в нашем динамически изменяющемся мире с высокой вероятностью, учитывая динамику каких-либо количественных показателей. Основными прогнозными значениями в этом направлении являются экспертные оценки.

Задача востребованности специальности со стороны абитуриентов также является сложной и неформализованной. Это обусловлено большим количеством качественных и количественных факторов влияющих на выбор абитуриентом той или иной специальности или ВУЗа. Учитывая различные типы факторов, неоднородность их влияния, сложность получения значений многих из них можно сделать вывод о невозможности применения линейных методов, приводящих их все к одному знаменателю. В данном случае целесообразно применение методик определяющих нелинейную зависимость общего результата от начальных данных. Такими инструментами являются нейронные сети и экспертные системы, но и эти методы имеют свои недостатки, поэтому использование только одного из методов не позволит достичь наилучшего результата. Для обработки одних данных будет целесообразно применить правила продукции, для других – ней-

ронные сети, для третьих будет достаточно формул.

Гибридные экспертные системы (ГЭС) – это такой метод решения неформализованных задач, с использованием которого задачу можно представить в виде иерархического дерева подзадач и указать для каждой из них свой метод решения []. В основу работы гибридных экспертных систем положена гибридная модель представления знаний. Она рассматривает задачу как совокупность подзадач образующих иерархию. Всем вершинам графа связей сопоставляются свои методы решения с индивидуальным набором параметров, который позволяет получить решения на всех уровнях иерархии.

В процессе прохождения через гибридную модель информация сжимается и, в конечном счете, сворачивается в единственный показатель – комплексную оценку. Однако, все значения промежуточных вершин являются доступными.

В общем виде процесс решения задач оценки в рамках гибридной экспертной системы представляет собой последовательное решение следующих задач:

1. Настройка гибридной экспертной системы, выбор метода решения для каждой подзадачи, формирование баз знаний для всех используемых методов по каждой сопоставленной им подзадаче).
2. Расчет прогнозного значения.
3. Интерпретация и объяснение полученной оценки.

На этапе настройки важной задачей является определение необходимого и достаточного количества входных параметров модели прогнозирования. Для этого использовался метод коллективных оценок руководства и специалистов организации (метод мозговой атаки); был определен следующий набор входных параметров:

- общее количество выпускников средних учебных заведений в текущем году;
- количество выпускников средних учебных заведений окончивших на хорошо и отлично;
- проходной балл при поступлении на специальность в предыдущем году;
- конкурс при поступлении на специальность в предыдущем году;

- количество мест занятых медалистами;
- клан набора на специальность в текущем году;
- стоимость обучения (при прогнозе спроса на платную образовательную услугу);
- доходы населения;

- престиж специальности;
- востребованность специальности работодателем;
- положение ВУЗа на рынке ОУ;
- степень удовлетворенности потребителей ОУ.

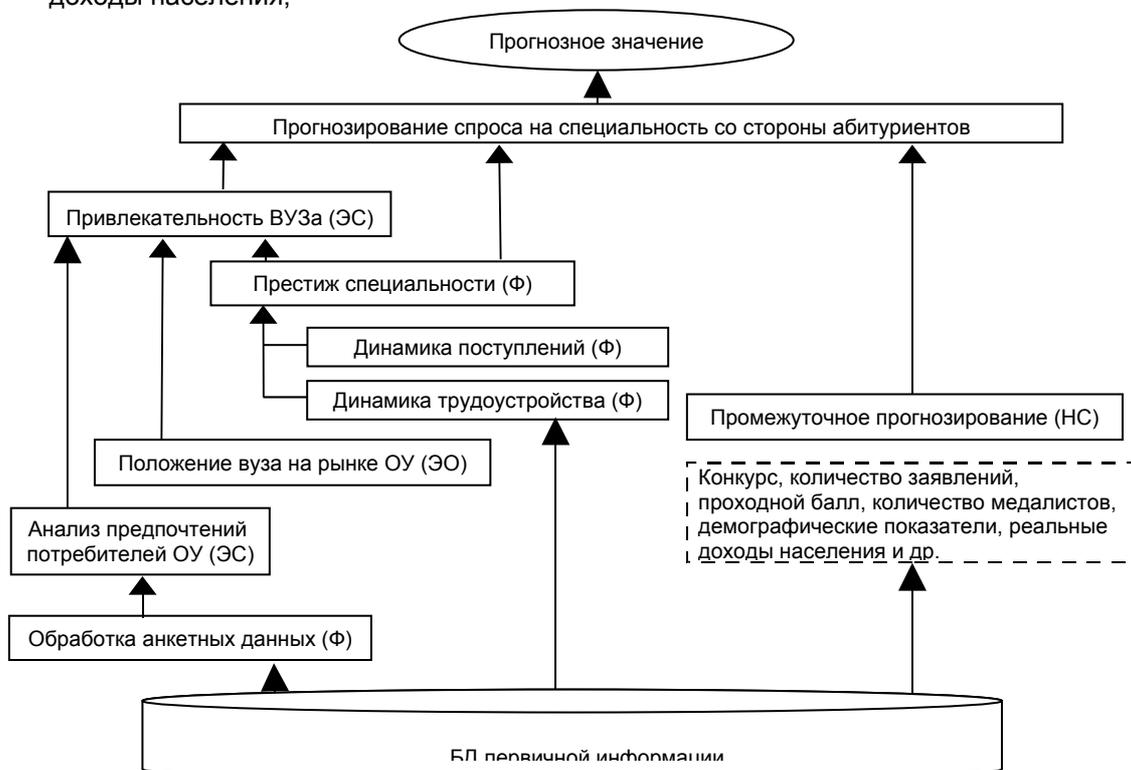


Рисунок 1 – Модель прогнозирования спроса

Список данных априорно является неполным, поскольку невозможно в общем случае описать все независимые и зависимые признаки, существенные для моделирования объекта или процесса. Это связано и с нашим ограниченным представлением о моделируемом объекте, и с ограничениями на возможность проведения тех или иных измерений.

В результате проблема выбора наиболее информативной подсистемы признаков приобретает важное значение, поскольку уменьшение числа признаков часто улучшает качество решения (и сокращает экономические и временные затраты на измерения или сбор информации). Желательно иметь возможность определения значимости каждого признака для принятия решения и выделения минимально необходимого набора базовых признаков для прогнозирования целевого признака с заданной точностью.

На рис. 1 представлена общая модель прогнозирования определенная экспертным путем. Модель не претендует на абсолютное

сходство с реальными процессами, а призвана продемонстрировать возможности использования гибридной экспертной системы. В качестве обучающей выборки были использованы данные за период с 1999 по 2004 гг. по 10 специальностям. Выборка не является достаточной, исследования, касающиеся малых выборок, рассмотрены в работах А.Н. Колмогорова, Дж. Неймана и А. Вальда. А.Н. Колмогоров установил критерий достаточности статистики при ограниченном числе наблюдений []. Учитывая специфику некоторых данных, их значения были сдвинуты на год вперед.

В контексте данной работы под нейросетью понимается сложная вектор-функция:

$$F^p(\mathbf{a}, \mathbf{x}) = \sum_{i_k=0}^{m_{k+1}} (a_{i_k, p}^{k+1} \cdot f_{i_k}^k(\mathbf{a}^k, \mathbf{f}^{k-1}(\dots, \mathbf{f}^2(\mathbf{a}^2, \mathbf{f}^1(\mathbf{a}^1, \mathbf{x}))))), \quad (1)$$

где p – номер компоненты выходного вектора, \mathbf{a} – вектор параметров или весов связей, \mathbf{x} – вектор входных данных или переменных, k – число слоев сети, m_k – число нейронов в k -м слое, $f_i^k(\mathbf{a}, \mathbf{f}^{k-1})$ – функция поведения нейрона

ПРОГНОЗИРОВАНИЕ ВОСТРЕБОВАННОСТИ СПЕЦИАЛЬНОСТИ С ИСПОЛЬЗОВАНИЕМ ГИБРИДНОЙ ЭКСПЕРТНОЙ СИСТЕМЫ

Для сравнения результатов прогнозирования нейронной сети с линейными методами была использована линейная аппроксимация по методу наименьших квадратов.

$$\varphi(x) = \sum_{k=1}^n a_k \varphi_k(x). \quad (2)$$

При эксперименте было обучено 45 нейронных сетей. Количество слоев в сети изменялось от 1 до 3, количество нейронов от 2 до 10. Сети были разбиты на три группы в зависимости от значения параметра сигмоиды. Анализ полученных результатов прогнозирования показал, что наилучшие значения были получены при значении параметра сигмоиды равном 1. При проведении эксперимента обнаружена следующая зависимость – с уменьшением количества нейронов в последнем слое уменьшалась средняя ошибка (рис. 2).

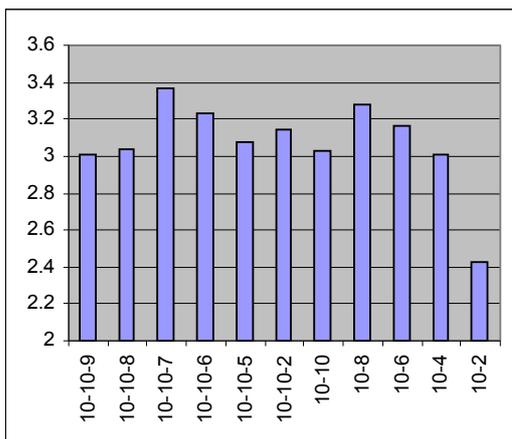


Рисунок 2 – Зависимость средней ошибки от количества нейронов в слое

В выбранной группе была определена сеть с наименьшей средней ошибкой при решении данной задачи. Оптимальная сеть состояла из одного слоя в 9 нейронов. Затем сеть была подвергнута контрастированию, после данной операции число нейронов сократилось до 5. Однако ошибка, получаемая при прогнозировании, была достаточно велика. На основании этого были сделаны выводы о нехватке набора входных параметров для правильного обучения сети.

В обучающую выборку были добавлены значения планового набора на бюджетные и внебюджетные места. После этого на основе новой выборки была создана сеть со структурой идентичной оптимальной сети полученной ранее. Ошибка по результатам, полученным после обучения сети, является допустимой и не превышает 9 %.

Первоначальное обучение было проведено по данным 1999-2000 гг. Средняя ошибка вычислений составила 8 %. Затем сеть была дообучена по данным 2001 г. После этого этапа средняя ошибка составила 6,5 %, и т.д. по всем годам выборки. Результаты обучения нейронных сетей приведены в таблице 1.

Таблица 1

Кол-во лет	Года	Ср.ошибка
2	1999, 2000	8,2 %
3	1999, 2000, 2001	6,8 %
4	1999, 2000, 2001, 2002	5,9 %
5	1999, 2000, 2001, 2002, 2003	8,9 %
6	1999, 2000, 2001, 2002, 2003, 2004	7,1 %

Как видно из представленных результатов, с увеличением обучающей информации значение ошибки уменьшалось. Однако с учётом 2003 г. ошибка сильно увеличилась, что дает основание говорить о появлении нового значимого фактора, не учтенного в нашей выборке. Этим фактором явился Единый государственный экзамен (ЕГЭ) и участие ВУЗа в эксперименте по ЕГЭ, что сократило общее количество абитуриентов и вызвало значительное увеличение ошибки.

Как видно из эксперимента, наличие в выборке одних статистических данных может привести к дестабилизации при появлении нового качественного фактора. Это еще раз подтверждает необходимость использования нескольких методов при решении сложных задач. После получения промежуточного значения на него были наложены экспертные оценки и правила продукции. Результаты приняты значения представленные в таблице 2.

Таблица 2

Кол-во лет	Года	Ср.ошибка
2	1999, 2000	7,8 %
3	1999, 2000, 2001	6,5 %
4	1999, 2000, 2001, 2002	5,4 %
5	1999 -2003	4,9 %
6	1999 - 2004	4,8 %

Для наглядности полученных экспериментов на рис. 2 приведены графики реального количества поданных заявлений, прогнозное значение на основе метода наименьших квадратов, график промежуточного

прогноза и график общего прогноза. Из представленных графиков видно, что наилучшие

результаты получены при использовании гибридной экспертной системы.

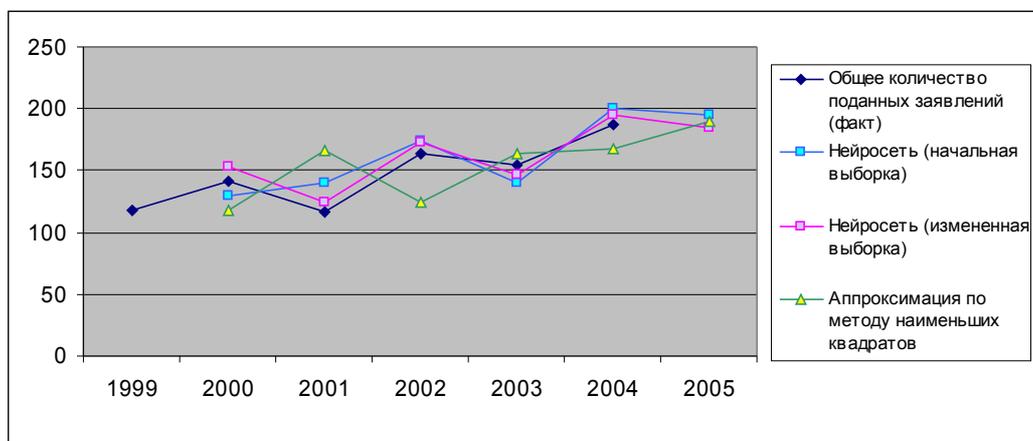


Рисунок 3 – Результаты прогнозирования спроса на специальность «Прикладная информатика в экономике»

В качестве вывода следует сказать, что применение интеллектуальных компонент в маркетинговых исследованиях образовательных услуг, в МИС, в частности нейронных сетей, имеет все основания. Однако необходимы достаточные обучающие выборки, определение основного круга факторов влияющих на результат прогноза и влияние результатов маркетинговых исследований на стратегическое планирование деятельности ВУЗа.

ЛИТЕРАТУРА

1. Пятковский О.И. Интеллектуальные компоненты автоматизированных информационных систем управления предприятием. Монография/ Барнаул: Изд-во АлтГТУ, 1999. – 355 с.
2. Колмогоров А.Н. Определение центра рассеивания и меры точности по ограниченному числу наблюдений. М: ИАН СССР, 1947.