

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.047

Малькевич
Егор Дмитриевич

Модели и алгоритмы краткосрочного прогнозирования уровня
заработной платы

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1-408005 – Математическое и программное
обеспечение вычислительных машин, комплексов и компьютерных сетей

Научный Руководитель
Глухова Лилия Александровна
Доцент, к.т.н

Минск 2014

КРАТКОЕ ВВЕДЕНИЕ

Постоянное изменение рынка труда, появление новых сфер занятости, а так же изменения в составе, структуре и качестве рабочей силы создают необходимость в подготовке технических специалистов, способных понимать и адаптировать свои знания и навыки для рынка. Актуальность темы диссертации заключается в предложении одного из возможных путей решения проблемы занятости путем анализа рынка труда по конкретным специальностям, используя методы краткосрочного прогнозирования.

В настоящее время в области краткосрочного прогнозирования уровня заработной платы на предприятиях и на рынке труда в целом, существуют проблемы, пути решения которых не разработаны. Например, происходит глобализация мировой экономики – преобразование мирового пространства в единую зону со свободным перемещением идей, информации, товаров, услуг и капитала. Сегодня отсутствует независимая система прогнозирования заработной платы, которая бы учитывала факторы и тенденции, влияющие на конкретную специальность. Это приводит к практически непредсказуемому формированию заработной платы, а также непредсказуемому спросу на специальности в системе образования.

Исследования, проводимые в данной области предлагают несколько различных подходов к решению данных проблем, в том числе и к прогнозированию изменений рынка труда.

Таким образом, предпосылкой для разработки темы данной магистерской диссертации стал тот факт, что не было найдено доступного и простого в использовании программного средства, обладающего необходимыми характеристиками.

Научная новизна данной работы заключается в предложении системы весов оценки влияния отдельно взятых параметров на итоговый прогноз. Это позволяет создать программное средство, способное адаптироваться под различные условия и изменения рынка труда. Таким образом, разработка системы краткосрочного прогнозирования, лишенной недостатков, связанных с точностью и полнотой получаемых данных, возможна.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цели и задачи исследования

Целью настоящей диссертационной работы является разработка научно обоснованных моделей и алгоритмов краткосрочного прогнозирования уровня заработной платы по специальности, позволяющих устранить или сократить недостатки существующих аналогов.

Достижение данной цели требует решения следующих задач:

- исследовать существующие модели и алгоритмы;
- разработать модель и алгоритм краткосрочного прогнозирования, способный учитывать множество параметров при составлении прогнозов;
- разработать алгоритм, способный учитывать достоверность получаемые данных;
- разработать модель и алгоритм для адаптации вычислений для отличающихся рыночных условий, а также для отличающихся специальностей;
- создание системы весов оценки получаемых параметров;
- провести сравнительный анализ для проверки гипотезы о повышенной точности разработанной модели и алгоритма краткосрочного прогнозирования заработной плат в сравнении с уже имеющимися программными решениями для выполнения аналогичных вычислений.

Объектом исследования являются заработная плата.

Предметом исследования являются краткосрочное прогнозирование уровня заработной платы.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность разработки качественного нового алгоритма прогнозирования, не требующего больших вычислительных мощностей, способного легко масштабироваться и адаптироваться под новые рынки и специальности в условиях реальной экономики.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий» по теме «Разработать модели, методы, алгоритмы для оценки параметров, повышения надежности и качества функционирования аппаратно-программных средств, систем и сетей сложной конфигурации и внедрить в

современные обучающие комплексы» (ГБ № 11-2004, № ГР 20111065, научный руководитель НИР – В. В. Бахтизин).

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично.

Вклад научного руководителя Л. А. Глуховой, заключается в формулировке целей и задач исследования.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена разработке моделей и алгоритмов для краткосрочного прогнозирования уровня заработной платы по специальности. В третьей главе предложена модульная архитектура приложения на основе весов оценки получаемых статистических данных, проведен сравнительный анализ существующего алгоритма и существующих аналогов.

Общий объем работы составляет 88 страниц, из которых основного текста 64 страниц, 31 рисунок на 16 страницах, 5 таблиц на 4 страницах, список использованных источников из 72 наименований на 4 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой** главе проведен анализ применяемых моделей, алгоритмов, используемых при решении задач краткосрочного прогнозирования уровня заработной платы.

Так же даны основные понятия и преодоления в области заработной платы.

Проведен анализ образцов открытого программного обеспечения для организации краткосрочных прогнозов по специальности: `salaris.dev.by`, `rabota.yandex.ru` и `payscale.com`. Выбор данного программного обеспечения был обусловлен такими критериями как возможность анализа заработной платы за период, прогнозирование заработной в зависимости от выбранного города, специальности и стажа.

Вторая глава посвящена разработке моделей и алгоритмов для краткосрочного прогнозирования уровня заработной платы по специальности.

Модели построены с помощью диаграммы связей, которые используются для создания, визуализации, структуризации и классификации идей, решения, принятия решений. Выделены основные факторы, влияющие на формирования заработной платы, с помощью диаграмм связей составлены модели, описывающие их взаимосвязь и влияние. На основе данных моделей составлены алгоритмы прогнозирования критериев.

На рисунке 1, показана общая модель влияния основных факторов с учетом «времени реакции» на формировании заработной платы, которая легла в основу построения моделей прогнозирования отдельных критериев и определению их влияние на заработную плату.

Выявлено, что непосредственно на заработную плату влияют такие факторы, как:

- законодательство;
- время и порог вхождения в профессию;
- внешние факторы, такие как: минимальная заработная плата за аналогичную работу в других странах, или на конкурентных предприятиях.

Для краткосрочного прогнозирования заработной платы было решено использовать факторы, оказывающие непосредственное влияние на конечную заработную плату.

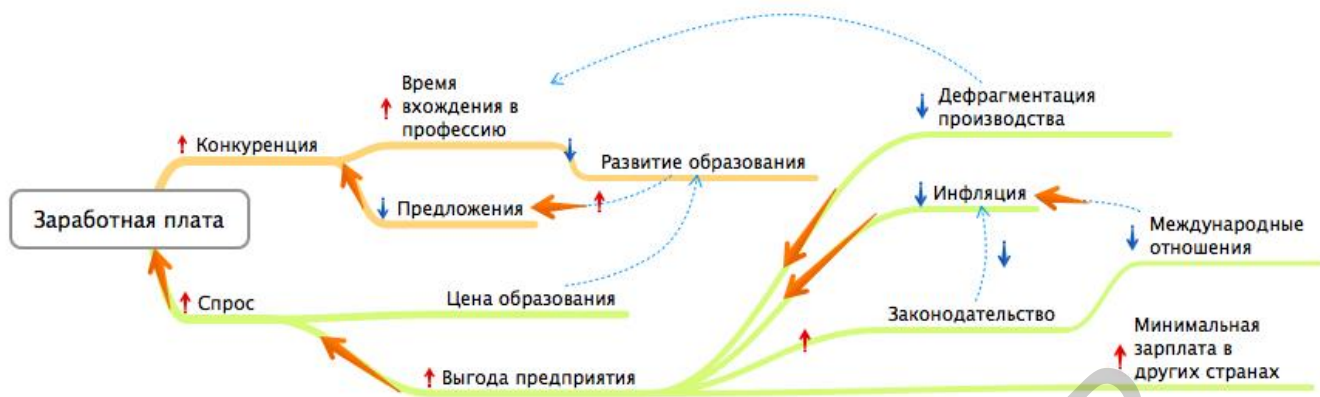


Рисунок 1 – Общая модель влияния основных факторов с учетом «времени реакции» на формирование заработной платы

Выявлено что, для разработки модели построения прогнозов, большинство факторы не подходят, так как их практически невозможно представить в числовом выражении, а также отслеживать изменения в режиме реального времени.

Для решения проблем зависимостей распределенных факторов и их структуризации выделен такой параметр как «время реакции».

Показано, что спрос является более инертным показателем, чем конкуренция. Это связано с тем, что он напрямую связан с потребностью предприятия в получении выгоды, при малейшем изменении выгод, под влиянием инфляции или дефрагментации производства.

Для расчета уровня изменения заработной платы необходимо учитывать множество факторов, которые оказывают либо прямое, либо косвенное влияние как друг на друга, так и на заработную плату в целом. Разработан обобщенный алгоритм расчета заработной платы, который изображен на рисунке 2.

Показано, что заработная плата находится в процессе непрерывного изменения, так же, как и колебания курсов валют на биржах. Анализ тенденций экономических изменений в стране, а также сам прогноз стали ключевыми параметрами в прогнозировании заработной платы.

Наличие большого количества параметров и коэффициентов, особенно достоверность, которых трудно определить на своем уровне, поставили под сомнение точность полученных результатов, и соответственно всего прогноза в целом.

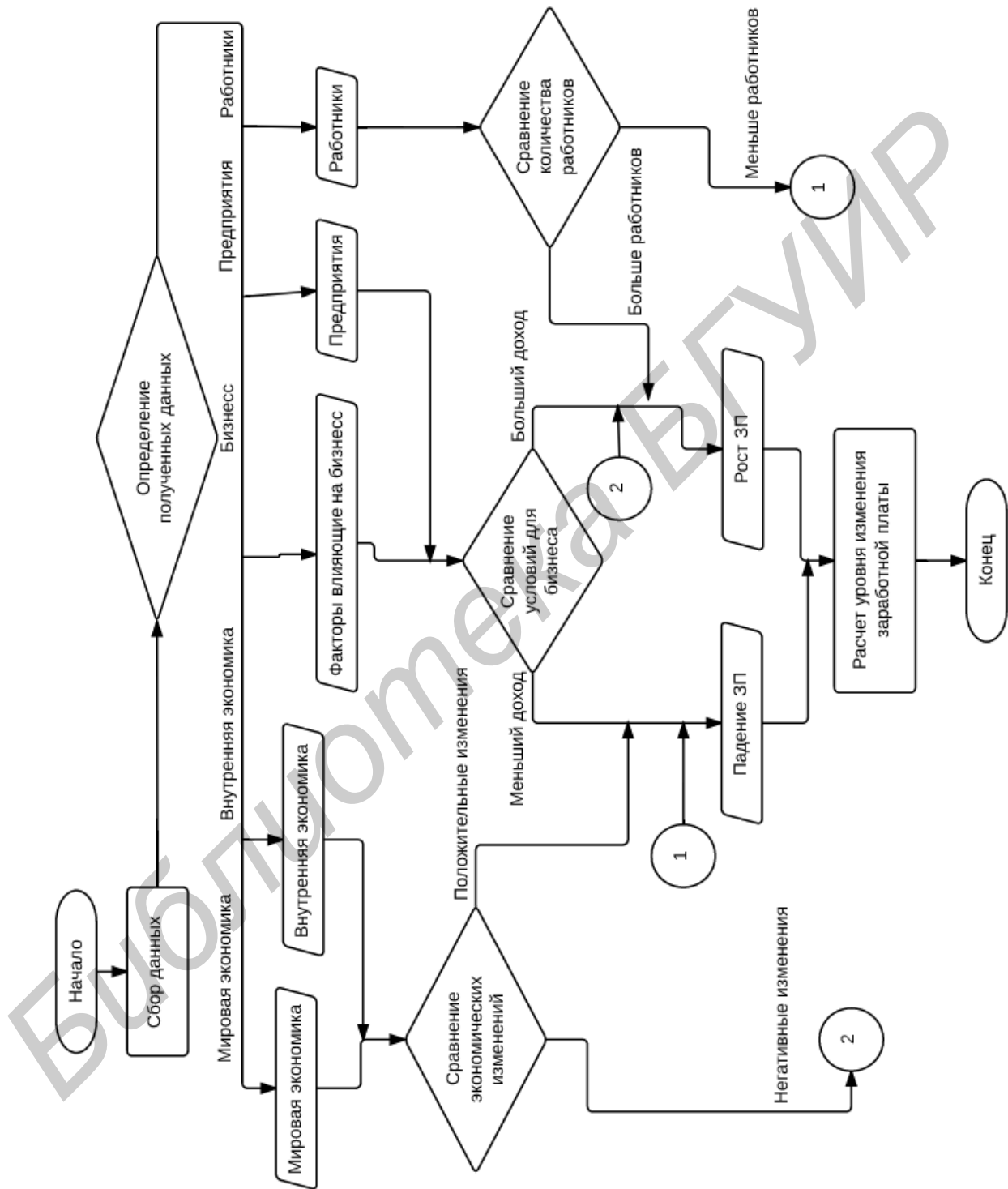


Рисунок 2 – Обобщенный алгоритм расчета заработной платы

Выявлено, что для анализа совокупности всех факторов, а так же составления прогнозов на их основе, обычных линейных алгоритмов недостаточно. Для поиска решений в многомерных пространствах поиска лучше всего подходят генетические алгоритмы. На практике получается очень точный инструмент прогнозирования, используя неполные или неоднородные статистические данные, без расчета фактов и взаимосвязей. На рисунке 3 показана схема работы генетического алгоритма.

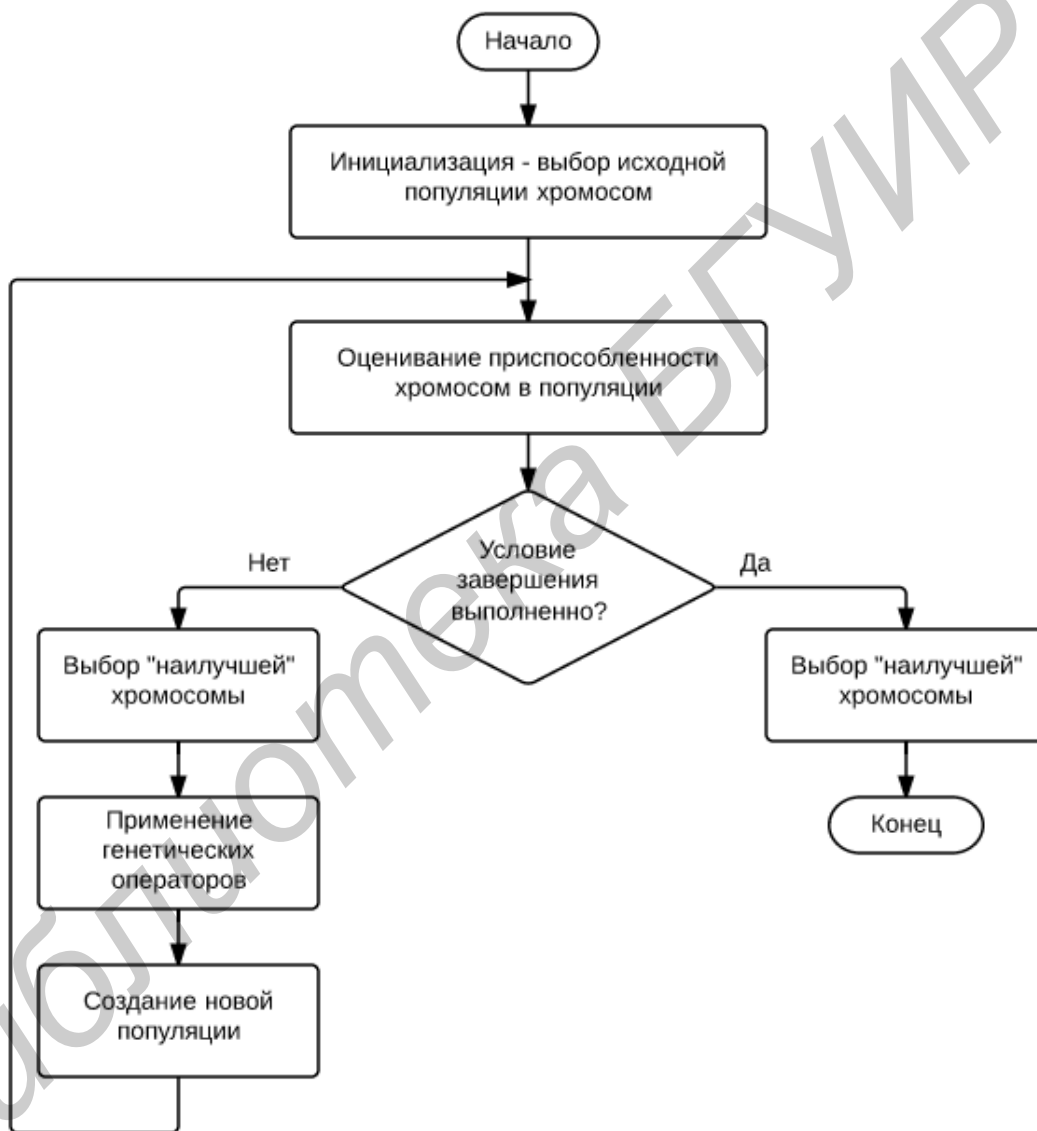


Рисунок 3– Схема работы генетического алгоритма

К сожалению, разработка алгоритма обучения генетическому алгоритму, позволяющему производить интервальные прогнозы по любым специальностям зависимых от региона, представляется невыполнимой задачей в рамках текущей работы. И очень сложной задачей вообще, так как с увеличением обрабатываемых параметров экспоненциально растет сложность обучения.

Лучшим подходом для решения тех же задач, но не используя генетические алгоритмы в полном объеме, выбран модульный подход. Для оценки влияния одних модулей на другие используется основной принцип генетических алгоритмов – система «весов влияния».

Таким образом разработан алгоритм работы одного модуля прогнозирования параметра, который показан на рисунке 4. Он не имеет зависимостей от других модулей.

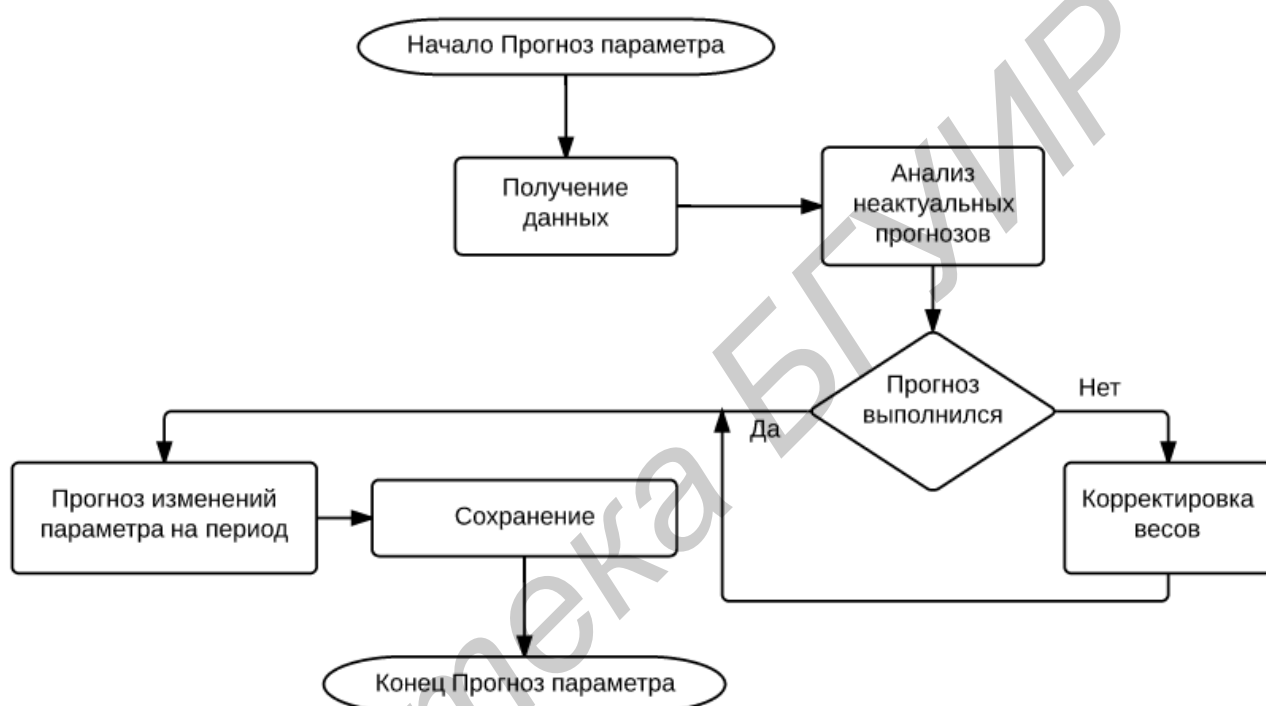


Рисунок 4 – Алгоритм одного модуля прогнозирования параметра

Важно следующее, для корректной работы алгоритма на вход подаются такие данные:

- вес или несколько весов;
- предыдущие прогнозы;
- массив параметров.

В **третьей** главе предложена архитектура построения приложения, учитывающая особенность корректировки весов на основе предыдущих прогнозов. Произведен анализ и последующее разделение общей архитектуры приложения на три основных модуля, что дало принципиально новые возможности для работы со статистическими данными.

Выделены общие черты всех модулей, разработанных во второй главе:

- сбор данных;
- анализ полученных данных;
- расчет тенденций;

- уточнение параметра важности, с учетом предыдущих прогнозов;
- сохранение данных.

На рисунке 5 показана схема данных всего приложения. Сбор и анализ полученных данных выделены в два больших процесса, которые выносятся в отдельные модули.

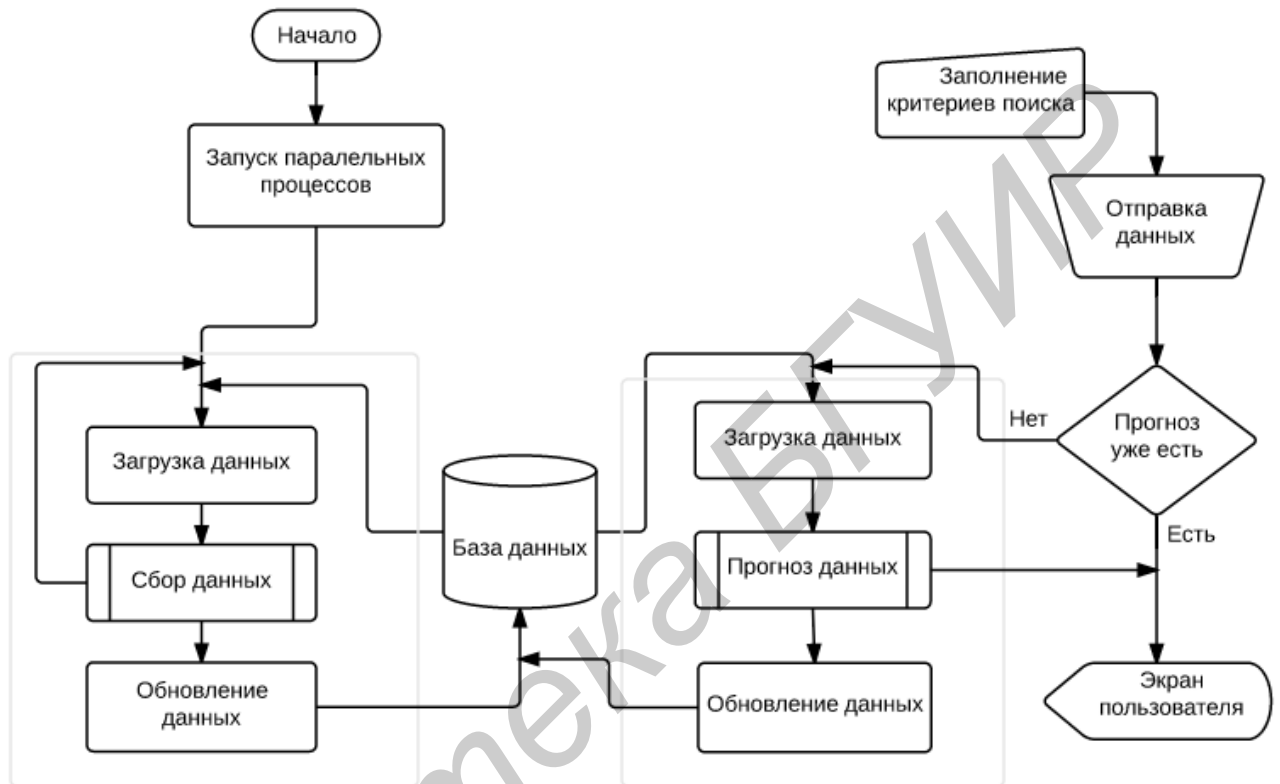


Рисунок 5 – Схема данных всего приложения

Сбор данных полностью независимый процесс, и может выполняться параллельно независимо от всего остального приложения.

Таким образом, получается независимый процесс, который запускает подпроцессы для получения данных по каждому из параметров из сети Интернет, нормализует их и сохраняют в базу данных для дальнейшего использования. Такая архитектура независима и легко масштабируется.

На основе имеющихся данных за 8 месяцев были составлены прогнозные модели с использованием программного средства, на основе предложенных алгоритмов, а также средствами rabota.yandex.ru и payscale.com. Прогнозы производились на месяц вперед. Сравнительная оценка погрешностей по каждому из прогнозов представлены в таблице 1.

По сравнению с прогнозами payscale.com и rabota.yandex.ru, где средняя погрешность находилась примерно в одном диапазоне, алгоритм продемонстрировал меньшую устойчивость, и большие колебания в прогнозах от 5.1% до 0%.

Таблица 1 – Сравнительная оценка погрешностей по каждому из прогнозов

	Август	Сентябрь	Октябрь
Разработанное приложение	+4.1%--+6.14% (+5.1%)	-4.4% – +1.2%(-1.6%)	-0.47%- +0.47% (0%)
Rabota.yandex.ru	-2.2%	-1.47%	- 0.3%
Payscale.com	-1.14%- +1.43% (+0.15%)	-0.87%-0.53% (-0.17%)	- 0.6% – +0.8%(+0.2%)
Реальные данные	100%	100%	100%

По сравнению с прогнозами payscale.com и rabota.yandex.ru, где средняя погрешность находилась примерно в одном диапазоне, алгоритм продемонстрировал меньшую устойчивость, и большие колебания в прогнозах от 5.1% до 0%.

К сожалению, такие колебания обусловлены алгоритмом обучения, которому данных за 7 месяцев оказалась недостаточно. Также сильные экономические колебания могут сбить веса, тем самым увеличив разброс вероятностных значений. При условии добавления большего количества модулей обработки данных, точность легко сможет превзойти имеющиеся на текущий момент программные средства, при условии, что экономические изменения не могут происходить стихийно, а происходят планомерно.

Важной особенностью разработанного алгоритма является то, что отслеживаются веса влияния одних параметров на другие. Тем самым возможно изучать реальные экономические отношения на рынке труда используя побочные данные программы.

При условии, что со временем некоторые важные параметры сегодня станут неактуальными в будущем, алгоритм способен показать устойчивость к изменениям.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Проведен анализ существующих средств и методов осуществления краткосрочного прогнозирования уровня заработной платы, который показал отсутствие объективного, надежного и простого в использовании алгоритма для осуществления точных прогнозов при использовании недостоверных сведений.
2. Сформулированы методы для работы с данными, необходимые для обеспечения целостности работы алгоритма.
3. Разработаны и исследованы алгоритмы построения прогнозов.
4. Разработан гибридный алгоритм прогнозирования уровня заработной платы.
5. Разработаны меры для обеспечения устойчивости алгоритма прогнозирования при отсутствии возможности объективно проверить достоверность данных.
6. Выявлены свойства, которые позволяют повысить эффективность прогнозирования уровня заработной платы.
7. Показана возможность реализации текущего алгоритма.
8. Произведена сравнительная оценка точности разработанного алгоритма и популярных систем для краткосрочного прогнозирования.
9. Показаны возможности различного применения результатов данной работы.

Рекомендации по практическому использованию результатов

В ходе анализа разработанных алгоритмов, выявлены свойства, которые позволяют повысить эффективность прогнозирования уровня заработной платы.

Программное средство построенное на основе разработанных моделей и алгоритмов, сможет осуществлять составления прогнозов существенно быстрее имеющихся аналогов, при условии сокращения вычислительных мощностей.

Использование однородные структур для работы с данными, дает возможность к простому горизонтальному масштабированию программных средств. Все это дает возможность работы с широким набором параметров отражающих рыночные отношения.

Список опубликованных работ

1-А. Малькевич Е.Д. Оценка влияния формы получения высшего образования на заработную плату специалиста / Е. Д. Малькевич, Л. А. Глухова // Дистанционное обучение - образовательная среда XXI века: Материалы VIII Междунар. науч.-метод. конференции (Минск, 5-6 декабря 2013г.). – Мн., БГУИР, 2013. – С. 88.

2-А. Малькевич Е. Д. Краткосрочное прогнозирование уровня зарплаты как фактор развития образования / Е. Д. Малькевич, Л. А. Глухова // Образование и наука в современных условиях : материалы II междунар. науч.–практ. конф. 2015 / редкол.: О. Н. Широков. – Чебоксары: ЦНС «Интерактив плюс», 2015. – ISBN 978-5-906626-56-1. С. [в публикации].

Библиотека БГУИР