УДК 004.41:004.891

Обзор алгоритмов прогнозирования трудоемкости для проектов на платформе «1С:Предприятие»

Н.А. Чичиль✉

Воронежский институт высоких технологий, Воронеж, Россия

Рассмотрены современные методы прогнозирования трудоемкости задач разработки и сопровождения решений на платформе «1С:Предприятие». На основе опубликованных исследований сравниваются регрессионные модели, методы на основе деревьев решений, ансамблевые алгоритмы и языковые модели, применяемые для оценки трудоемкости разработки программного обеспечения. Отдельно рассматриваются данные, используемые при построении прогноза, включая структурированные характеристики задач, текстовые описания обращений, экспертные оценки и исторические данные.

Ключевые слова: прогнозирование трудоемкости, машинное обучение, ансамблевые методы, языковые модели, экспертная оценка, «1С:Предприятие».

Review of Software Effort Estimation Methods for Projects on the 1C:Enterprise Platform

N.A. Chichil✉

Voronezh Institute of High Technologies, Voronezh, Russia

This paper reviews contemporary methods for software effort estimation in the development and maintenance of solutions based on the 1C:Enterprise platform. Published studies are analyzed to compare regression models, decision tree-based methods, ensemble learning algorithms, and language models applied to software effort estimation. Particular attention is paid to the types of data used for prediction, including structured task attributes, textual issue descriptions, expert estimates, and historical data.

Keywords: software effort estimation, machine learning, ensemble learning methods, language models, expert estimation, 1C:Enterprise.

Введение

При разработке и сопровождении решений на платформе «1С:Предприятие» оценка трудоемкости используется при планировании сроков выполнения работ, распределении нагрузки между исполнителями и определении стоимости проектов. Во многих организациях она по-прежнему основывается на экспертной оценке. А.В. Мельников, Ф.А. Музалевский и Р.А. Солодуха рассматривают экспертную оценку как один из основных элементов определения трудоемкости разработки на платформе «1С:Предприятие» [1].

В последние годы опубликован ряд исследований, посвященных прогнозированию трудоемкости разработки программного обеспечения. Banimustafa рассматривает регрессионные модели, деревья решений, ансамблевые методы и нейронные сети [2]. Satapathy отмечает, что ни один метод не показывает лучшие результаты во всех случаях: итог зависит от свойств выборки, уровня шума и характера прогнозируемой зависимости [3].

При сопровождении решений на платформе «1С:Предприятие» в информационной системе обычно хранится больше сведений, чем используется в большинстве исследований Software Effort Estimation. Помимо структурированных характеристик задачи сохраняются текстовые описания обращений, экспертные оценки, сведения об исполнителях и история выполненных доработок. Часть этой информации уже представлена в виде отдельных признаков. Остальная остается только в тексте.

Одни методы используют только структурированные данные. Другие работают с текстовыми описаниями задач. Комбинированные подходы объединяют оба источника информации [5, 6]. Различия между такими подходами связаны не только с используемыми алгоритмами, но и с данными, на которых строится прогноз.

Особенности применения методов прогнозирования для задач на платформе «1С:Предприятие»

В большинстве исследований по Software Effort Estimation прогноз строится на основе структурированных характеристик проекта [2, 3]. Обычно используются размер программного продукта, численность команды, продолжительность разработки и другие количественные показатели. Такой набор данных остается основой большинства исследований и позволяет сравнивать методы в одинаковых условиях.

В методике А.В. Мельникова, Ф.А. Музалевского и Р.А. Солодухи используются тип и количество объектов конфигурации, объем программного кода и экспертные оценки трудоемкости [1]. Помимо характеристик задачи сохраняются сведения о проекте, исполнителях, экспертных оценках и фактической трудоемкости завершенных работ. В информационной системе одновременно хранятся текстовые описания обращений, комментарии специалистов и история выполненных доработок.

Структурированных признаков недостаточно. Количество изменяемых объектов, категория обращения и приоритет дают только общее представление о задаче. Детали реализации в этих признаках не отражаются.

Две задачи могут иметь одинаковую категорию обращения, одинаковый приоритет и одинаковое количество изменяемых объектов. В одном случае требуется добавить реквизит в существующую форму документа. В другом – изменить механизм расчета, используемый несколькими подсистемами. Формальные характеристики почти совпадают. Трудоемкость – нет.

Описание обращения обычно содержит сведения, которых нет среди структурированных признаков задачи. Именно на такую информацию опирается эксперт при оценке новой доработки.

Специалист сопоставляет не только категорию обращения или изменяемые объекты. Он учитывает содержание задачи, особенности реализации и результаты похожих доработок.

После завершения проекта в информационной системе сохраняются описание обращения, фактическая трудоемкость, сведения об исполнителе, проекте и экспертной оценке. Эти данные могут использоваться повторно. А.Д. Ковалев рассматривает векторное представление текстов и поиск семантически близких запросов как средства автоматизации сопровождения программного обеспечения [4].

При таком составе данных сравниваются уже не только алгоритмы. Не меньшее значение имеет набор сведений, который используется при построении прогноза.

Сравнение методов прогнозирования трудоемкости

Для оценки трудоемкости разработки программного обеспечения используются различные методы прогнозирования [2, 3]. Они различаются не только используемыми алгоритмами, но и составом исходных данных. Одни методы работают только со структурированными признаками. Другие используют текстовые описания задач. Есть и подходы, объединяющие оба источника информации [2, 3, 5, 6].

Для задач разработки и сопровождения решений на платформе «1С:Предприятие» различия между этими подходами проявляются особенно отчетливо. В информационной системе одновременно хранятся сведения о проекте, категории обращения, исполнителе, экспертной оценке, текстовые описания задач, комментарии разработчиков и история выполненных доработок. Часть информации хранится в виде отдельных признаков. Остальная содержится только в текстовом описании обращения.

Используемые источники данных в рассмотренных публикациях приведены в таблице 1.

Таблица 1

Использование различных источников данных методами прогнозирования

Источник данных	Регрессионные модели	Деревья решений	Бустинг	Языковые модели	Комбинированные методы
Структурированные признаки	+	+	+	–	+
Текстовое описание	–	–	–	+	+
Экспертная оценка	±	+	+	±	+
Непосредственное использование аналогичных ранее выполненных задач	–	–	–	±	+

Во всех рассмотренных работах подготовка исходной информации рассматривается как обязательный этап построения модели независимо от выбранного алгоритма.

Регрессионные модели остаются одной из наиболее распространенных отправных точек при оценке трудоемкости [2, 3]. Их часто используют как базовый вариант при сравнении различных методов. Модель использует только структурированные признаки.

Random Forest и Extra Trees используют деревья решений как основу модели [2, 3]. По сравнению с регрессионными моделями они учитывают более сложные сочетания признаков.

LightGBM относится к алгоритмам градиентного бустинга на деревьях решений. Ансамбль строится последовательно: каждое новое дерево уточняет ошибки, оставшиеся после предыдущих итераций. Для ускорения обучения в LightGBM применяются односторонняя выборка на основе градиента и объединение взаимоисключающих признаков. В экспериментах авторов эти механизмы сократили время обучения при близкой точности на крупных наборах данных [7].

Все перечисленные методы используют заранее подготовленные признаки. Если часть информации содержится только в текстовом описании обращения, эти сведения не участвуют в построении прогноза.

В работе GPT2SP для построения прогноза используется текстовое описание задачи [5]. Текстовое описание задачи преобразуется в векторное представление,

которое затем используется при оценке трудоемкости [5]. GPT2SP не ограничивается получением числовой оценки. Разработанный авторами прототип выделяет слова, повлиявшие на прогноз, и показывает примеры ранее оцененных задач из обучающей выборки [5].

Текстовое описание задачи обычно не содержит сведений о проекте, опыте исполнителя или результатах предварительной экспертной оценки. Такие сведения уже содержатся в корпоративной информационной системе и могут использоваться при прогнозировании трудоемкости.

Fine-SE объединяет семантические признаки текста с экспертными признаками, сформированными по результатам интервью с 26 сотрудниками предприятия [6]. Модель проверялась более чем на 30 000 задачах из промышленных и открытых проектов. В прогнозе одновременно используются текстовое описание обращения и характеристики проекта.

Различия между рассмотренными методами приведены в таблице 2.

Таблица 2

Сравнение методов прогнозирования по используемым данным

Метод	Особенность применения	Типичные условия использования
Регрессионные модели	Зависимость задается через подготовленные признаки	Небольшие и однородные выборки
Random Forest / Extra Trees	Учитывают сложные сочетания признаков	Разнородные табличные данные
LightGBM	Градиентный бустинг с оптимизацией обработки наблюдений и признаков	Большие табличные выборки
Языковые модели	Анализируют текст задачи	Информативные текстовые описания
Комбинированные методы	Объединяют текстовые и структурированные сведения	Корпоративные системы с несколькими источниками данных

В ранних работах обычно сравниваются регрессионные модели, деревья решений, нейронные сети и ансамблевые методы [2, 3]. Главная задача – выбрать модель, которая точнее прогнозирует трудоемкость. Набор исходных данных при этом почти не обсуждается.

Позднее появляются работы, где внимание переносится на другую сторону задачи. Авторы начинают исследовать, какую информацию можно использовать при построении прогноза. В прогнозировании начинают использовать текстовые описания задач [5].

Появление LightGBM показывает еще одно направление развития ансамблевых методов – сокращение времени обучения и затрат памяти при работе с крупными и высокоразмерными наборами данных [7].

В более новых публикациях исследователи все чаще объединяют несколько подходов. Fine-SE показывает переход от использования одного типа признаков к объединению информации разной природы [6]. Вместе используются структурированные признаки и текстовое описание задачи. В работах GPT2SP и Fine-SE различается не только архитектура моделей, но и состав информации, используемой при прогнозировании [5, 6].

Условия применения методов в рассмотренных работах различаются. GPT2SP дополняет прогноз объяснением и примерами прошлых оценок [5], тогда как Fine-SE объединяет семантические и экспертные признаки [6]. Регрессионные модели и деревья решений при этом сохраняются как базовые методы сравнения [2, 3].

Заключение

Рассмотренные методы различаются алгоритмами и составом исходных данных. Регрессионные модели, деревья решений и бустинг работают с заранее подготовленными признаками. Языковые модели используют текст задачи, а комбинированные подходы соединяют оба источника информации [2, 3, 5, 6].

В проектах на платформе «1С:Предприятие» сведения о задаче представлены в разных формах. Это приходится учитывать при выборе способа прогнозирования трудоемкости.

СПИСОК ИСТОЧНИКОВ

1. Мельников А.В. Методика оценки трудоемкости разработки на платформе «1С:Предприятие 8» / А.В. Мельников, Ф.А. Музалевский, Р.А. Солодуха // Судебная экспертиза. – 2026. – № 1 (85). – С. 124–137.

2. BaniMustafa A. Predicting Software Effort Estimation Using Machine Learning Techniques / A. BaniMustafa // Proceedings of the 8th International Conference on Computer Science and Information Technology (CSIT 2018). – IEEE, 2018. – P. 249–256.

3. Satapathy Sh.M. Effort Estimation Methods in Software Development Using Machine Learning Algorithms / Sh.M. Satapathy // Academia.edu [Электронный ресурс]. – URL: https://www.academia.edu/71946155/Effort_Estimation_Methods_in_Software_Development_Using_Machine_Learning_Algorithms (дата обращения: 20.07.2026).

4. Ковалев А.Д. Методы автоматизации сопровождения программного обеспечения на основе интеллектуальной обработки запросов заказчика: автореф. дис. ... канд. техн. наук: 2.3.5 / Ковалев Артем Дмитриевич; Санкт-Петербургский политехнический университет Петра Великого. – Санкт-Петербург, 2023. – 18 с.

5. Fu M. GPT2SP: A Transformer-Based Agile Story Point Estimation Approach / M. Fu, Ch. Tantithamthavorn // IEEE Transactions on Software Engineering. – 2023. – Vol. 49, No. 2. – P. 611–625.

6. Fine-SE: Integrating Semantic Features and Expert Features for Software Effort Estimation / Y. Li, Zh. Ren, Zh. Wang [et al.] // Proceedings of the 46th IEEE/ACM International Conference on Software Engineering (ICSE 2024). – New York: ACM, 2024. – URL: https://ieeexplore.ieee.org/document/10548937 (дата обращения: 20.07.2026).

7. LightGBM: A Highly Efficient Gradient Boosting Decision Tree / G. Ke, Q. Meng, Th. Finley [et al.] // Advances in Neural Information Processing Systems 30. – 2017. – P. 3146–3154.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Чичиль Николай Анатольевич, студент, Воронежский институт высоких технологий, Воронеж, Россия.
e-mail: lord-nick@rambler.ru