

УДК 004.41:004.891

Интеллектуальное прогнозирование трудоемкости IT-задач на основе исторических данных платформы «1С:Предприятие» с использованием гибридного ансамблевого метода

Н.А. Чичиль✉

Воронежский институт высоких технологий, Воронеж, Россия

Предложен гибридный метод прогнозирования трудоемкости IT-задач при разработке и сопровождении решений на платформе «1С:Предприятие». Метод объединяет экспертную оценку, структурированные характеристики задачи, поиск семантически близких обращений по историческим данным и ансамбль моделей машинного обучения. Для формирования итогового прогноза используется метамодель, объединяющая результаты базовых алгоритмов и дополнительные признаки, полученные на основе исторических данных. Экспериментальная проверка выполнена на выборке, включающей более 12 тыс. завершенных задач. Предложенный метод обеспечил более высокую точность прогнозирования по сравнению с отдельными моделями машинного обучения и подтвердил эффективность совместного использования экспертных знаний, семантического поиска и исторических данных.

Ключевые слова: прогнозирование трудоемкости, машинное обучение, ансамблевое обучение, семантический поиск, экспертная оценка, 1С:Предприятие.

Hybrid Ensemble-Based Effort Prediction for IT Tasks Using Historical Data from the 1C:Enterprise Platform

N.A. Chichil✉

Voronezh Institute of High Technologies, Voronezh, Russia

This paper proposes a hybrid method for predicting the effort required for IT tasks related to the development and maintenance of solutions based on the 1C:Enterprise platform. The method combines expert estimation, structured task attributes, search for semantically similar references using historical data, and an ensemble of machine learning models. A meta-model is used to generate the final prediction by integrating the outputs of the base algorithms with additional features derived from historical data. The experimental verification was carried out on a sample that included more than 12,000 completed tasks. The proposed method demonstrated higher prediction accuracy than individual machine learning models and confirm the effectiveness of combining expert knowledge, semantic retrieval, and historical project data.

Keywords: effort prediction, machine learning, ensemble learning, semantic retrieval, expert estimation, 1C:Enterprise.

Введение

Оценка трудоемкости используется при планировании сроков выполнения работ, распределении нагрузки между исполнителями и расчете стоимости проектов. Для задач сопровождения решений на платформе «1С:Предприятие» она чаще всего выполняется экспертным путем и зависит как от опыта специалиста, так и от доступной информации о ранее выполненных работах.

Существующие методы оценки различаются набором исходных данных и способом формирования итоговой оценки. В работе А.В. Мельникова, Ф.А. Музалевского и Р.А. Солодухи приведен обзор наиболее распространенных подходов и отмечено, что большинство из них в той или иной степени опирается на экспертное оценивание [1]. Для разработки на платформе «1С:Предприятие 8» авторами предложена методика, основанная на сочетании экспертных оценок и метода PERT [1].

По мере накопления исторических данных при оценке становится возможным использовать не только характеристики новой задачи, но и сведения о ранее выполненных работах. В автореферате А.Д. Ковалева одним из направлений автоматизации сопровождения программного обеспечения рассматривается поиск семантически похожих запросов заказчиков [2]. Такой подход предполагает использование накопленной информации при обработке новых обращений [2].

В современных моделях прогнозирования трудоемкости текстовое описание задачи используется как самостоятельный источник информации. В модели GPT2SP для этой цели применяется трансформерная архитектура [3]. В работе Fine-SE семантические признаки объединяются с экспертными характеристиками задачи в рамках единой модели прогнозирования [4].

Семантический поиск позволяет находить близкие по содержанию обращения, экспертные признаки отражают особенности процесса разработки, а структурированные характеристики описывают формальные параметры задачи. Объединение этих источников информации в рамках единой модели прогнозирования представляет практический интерес для задач сопровождения решений на платформе «1С:Предприятие».

В работе исследуется возможность объединения экспертной оценки, структурированных характеристик задачи и исторических данных при прогнозировании трудоемкости ИТ-задач на платформе «1С:Предприятие».

Метод прогнозирования трудоемкости

Во многих моделях исторические данные используются только на этапе обучения. После построения модели прогноз рассчитывается по признакам новой задачи, тогда как сведения о ранее выполненных работах непосредственно в расчете уже не участвуют.

При оценке новой задачи специалисты обычно обращаются к ранее выполненным работам, сопоставляя их содержание и фактическую трудоемкость. Такая информация уже содержится в учетной системе и может использоваться непосредственно при прогнозировании. Поэтому при прогнозировании используется поиск аналогичных обращений.

Общая схема метода представлена на рисунке.

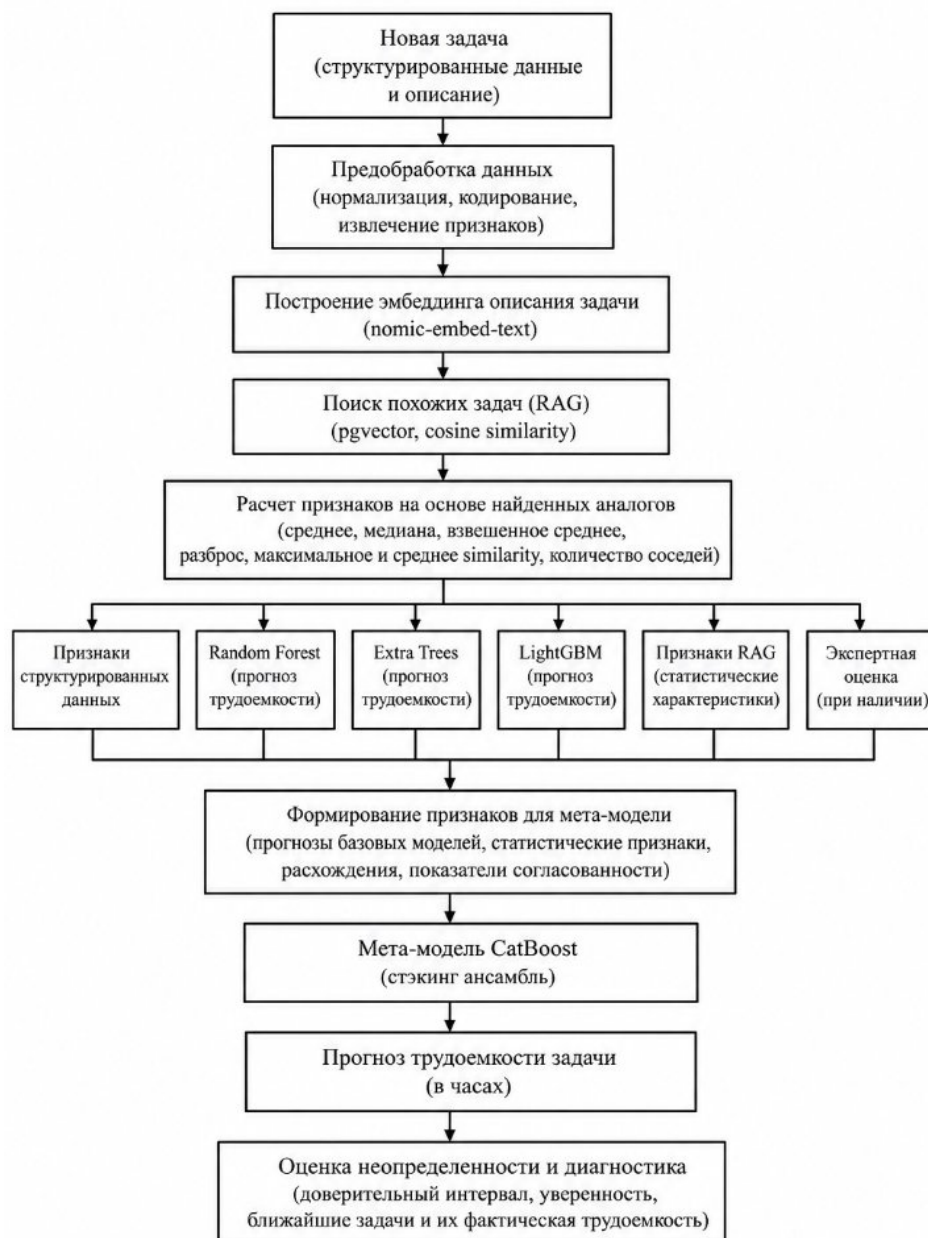


Рисунок. Общая схема метода

Результаты проверки показали, что использование только семантического поиска не обеспечивает требуемой точности прогнозирования. Семантически близкие задачи далеко не всегда имеют сопоставимую трудоемкость. Даже при практически одинаковом описании объем работ может отличаться в несколько раз, если затрагиваются разные подсистемы, используются различные механизмы платформы или требуется изменение большого числа объектов конфигурации. Само текстовое описание не содержит всей информации, определяющей трудоемкость.

По той же причине нельзя ограничиться и структурированными признаками. Они хорошо описывают формальные характеристики задачи, но практически ничего не говорят о содержании доработки. На выборке встречалось множество обращений с одинаковыми значениями категории, сложности и приоритета, фактическая трудоемкость которых существенно различалась.

Поэтому прогноз строится на основе нескольких независимых источников информации.

Несколько базовых моделей используются для того, чтобы учитывать различные зависимости между признаками задачи и фактической трудоемкостью.

Каждая из базовых моделей формирует собственный прогноз, опираясь на различные закономерности обучающей выборки. Это уменьшает влияние ошибок отдельных алгоритмов и делает итоговую оценку менее чувствительной к особенностям конкретной задачи.

Структурированные признаки поступают в Random Forest, Extra Trees и LightGBM. Эти модели обучаются независимо и формируют собственные оценки. Отдельно обрабатывается текст обращения. После построения эмбединга выполняется поиск наиболее похожих завершенных задач, после чего рассчитываются характеристики найденной группы: медиана, среднее значение, разброс трудоемкости, максимальное и среднее семантическое сходство. Если перед прогнозированием выполнена экспертная оценка, она также включается в расчет.

Признаки, полученные в результате семантического поиска, не используются как самостоятельный прогноз. Они дополняют структурированные характеристики задачи информацией о найденных аналогах, позволяя учитывать фактическую трудоемкость близких по содержанию обращений и степень их семантического сходства.

Все полученные признаки объединяются метамоделью CatBoost. Она получает не только прогнозы отдельных алгоритмов, но и информацию о том, насколько эти прогнозы согласуются между собой. В результате вклад каждого источника определяется по историческим данным, а не задается заранее.

Экспериментальное исследование

Проверка метода выполнялась на исторических данных промышленной системы сопровождения решений на платформе «1С:Предприятие». В выборку вошла информация о 12 451 завершенной задаче. Для каждой записи использовались фактическая трудоемкость, структурированные характеристики обращения и текстовое описание.

Данные были разделены на обучающую и тестовую выборки в соотношении 75:25. В обучающую выборку вошли 9338 задач, в тестовую – 3113. Качество прогнозирования оценивалось по метрикам MAE, RMSE, MAPE и коэффициенту детерминации R^2 .

В качестве базовых моделей использовались SVR, Random Forest, Extra Trees, GRNN и LightGBM. Результаты их работы сравнивались с предложенным методом и экспертной оценкой.

Результаты эксперимента приведены в таблице 1.

Таблица 1

Сравнение качества прогнозирования

Метод	MAE	RMSE	MAPE	R^2
Экспертная оценка	740,39	2717,44	9,28	0,946
SVR	3506,73	6196,46	59,83	0,722
Random Forest	2759,30	4071,11	51,38	0,880
Extra Trees	2737,70	4051,20	52,79	0,881
GRNN	5814,03	9450,99	98,72	0,352

Таблица 1 (продолжение)

Метод	MAE	RMSE	MAPE	R ²
LightGBM	2895,56	4364,07	51,52	0,862
Предлагаемый метод	979,07	2796,19	11,40	0,944

По данным таблицы 1 предложенный метод превосходит отдельные модели машинного обучения. При этом минимальная ошибка остается у экспертной оценки. Такой результат ожидаем, поскольку она формируется специалистом, знакомым с предметной областью и особенностями конкретной задачи. Вместе с тем экспертная оценка доступна не всегда. В таких случаях предложенный метод позволяет получить прогноз, близкий к экспертному, используя только исторические данные и описание обращения.

Для оценки вклада отдельных компонентов метода был выполнен абляционный анализ. Из модели последовательно исключались экспертная оценка, семантический поиск и отдельные базовые алгоритмы. Результаты приведены в таблице 2.

Таблица 2

Абляционный анализ

Конфигурация	MAE	Изменение
Полная модель	996,27	–
Без экспертной оценки	2681,40	+169,1%
Без семантического поиска	1112,18	+11,6%
Без Random Forest	943,14	–5,3%
Без Extra Trees	1029,10	+3,3%

Увеличение ошибки наблюдается при исключении экспертной оценки. Это подтверждает, что она остается наиболее информативным источником данных. Исключение семантического поиска также приводит к росту ошибки, хотя эффект выражен слабее. Влияние отдельных базовых моделей значительно меньше. Более того, исключение Random Forest привело к небольшому снижению MAE, что свидетельствует о частичном дублировании информации другими компонентами ансамбля.

Распределение вклада признаков подтверждается результатами SHAP-анализа. Наибольшее влияние на итоговый прогноз оказывает экспертная оценка, далее следуют прогноз Random Forest и признаки, характеризующие согласованность отдельных моделей. Признаки, полученные при семантическом поиске, вносят меньший вклад по отдельности, но в совокупности повышают устойчивость метода.

Заключение

Полученные результаты показывают, что использование только структурированных признаков или только семантического поиска не обеспечивает устойчивого качества прогнозирования. Лучший результат достигается при объединении нескольких независимых источников информации.

Экспертная оценка остается наиболее информативным признаком, но ее недостаточно для всей выборки задач. Использование исторических данных, найденных аналогов и ансамбля моделей позволяет уменьшить ошибку прогнозирования и сделать оценку более устойчивой.

Предложенный метод не заменяет экспертную оценку, а дополняет ее использованием исторических данных организации. Это позволяет получать прогноз трудоемкости новых задач на основе информации о ранее выполненных работах и может использоваться в существующем процессе сопровождения решений на платформе «1С:Предприятие».

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. 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.
3. 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.
4. 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: 18.07.2026).
5. Analysis of Software Effort Estimation by Machine Learning Techniques / M.U. Nisa, M. Saqlain, M. Abid [et al.] // Ingénierie des Systèmes d'Information. – 2023. – Vol. 28, No. 6. – P. 1445–1457.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Чичиль Николай Анатольевич, студент, Воронежский институт высоких технологий, Воронеж, Россия.

e-mail: lord-nick@rambler.ru