

Оценка возможности прогнозирования характеристик заторов льда на основе методов машинного обучения (на примере реки Сухона)

Цуй Юйсюань

Студент (бакалавр)

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: 1606398766@qq.com

Для рек севера Европейской части России, во время весеннего половодья часто образуются заторы льда, являясь весьма опасным природным явлением. Из-за большого числа факторов, влияющих на формирование заторов льда, часто при экстремальных погодных условиях, прогнозирование заторов льда затруднено. В России один из крупнейших специалистов в этой области S.Beltaos[1] разработал модель разрушения и движения ледяного покрова при вскрытии рек. В Китае прогнозирование заторов в последние годы основывается на методах машинного обучения. В США D.Massie и др. [3] использовали модели нейронных сетей для прогнозирования ледяных заторов на р. Аллегени, штат Пенсильвания. Целью данной работы является разработка методики прогноза возникновения ледового затора на р. Сухона при слиянии ее с р.Юг на основе методов машинного обучения (KNN-метода (K Nearest Neighbor) и дерева решений) и статистического моделирования (модель множественной линейной регрессии). Исследуемый участок реки расположен в низовьях р. Сухона, к северу от д. Каликино и к югу от г.Котлас, он один из наиболее проблемных с точки зрения заторных наводнений в этом бассейне.

В этой работе данные с 2000 по 2020 гг. обозначены как тестовый набор, а обучающий набор состоит из данных с 1960 по 1999 гг. В качестве целевой переменной (target) была выбрана величина, принимающая два значения: 0 (затора не было) и 1 (затор был). С помощью WoE-анализа (Weight of Evidence [2]) было отобрано 10 признаков - суммарное количество твердых осадков, максимальная толщина льда, уровень воды в период установления ледостава, особенности температурного режима в период замерзания и другие, которые в дальнейшем использовались при построении модели.

В машинном обучении использовались метод KNN и метод дерева решений соответственно (тестовый набор с 2000 по 2020 г., обучающий набор с 1960 по 1999 г.). Мы получили наилучшие результаты для модели KNN с метрикой Минковского, с пятью соседями, ассигасу=0.81, recall=1, что означает, что модель правильно определила все заторы. На основе модели множественной линейной регрессии для максимального заторного уровня воды и его заторной составляющей были получены факторы с наибольшей корреляцией: уровень воды в период установления ледостава, продолжительность осеннего ледохода и замерзания, максимальная толщина льда и расход воды до вскрытия, однако необходимы дальнейшие исследования для улучшения результатов.

Источники и литература

- 1) Beltaos S. A conceptual model of river ice breakup // Canadian Journal of Civil Engineering. 1984. Pp.516-529. DOI:10.1139/I84-071.
- 2) Weed D.L. Weight of Evidence: A Review of Concept And Methods // Risk Analysis. 2005. Vol. 25, No 6, p. 1545-1557.
- 3) Massie D. D., White K. D., Daly S. F., McDonald R. Predicting Ice Jams With Neural Networks // 21st International Conference on Offshore Mechanics and Arctic Engineering, 2002. Volume 3. doi:10.1115/omae2002-28551