Исследование возможности использования методов машинного обучения для автоматической классификации треков спутникового позиционирования на примере системы мониторинга рыболовства

А.Ю. Дегай[1], В.Н. Пырков[1], П.М. Василец[1], М.В. Андреев[1]

[1] Институт космических исследований РАН, Москва, e-mail: vpyrkov@mail.ru

Важность достоверности отчетов о вылове в отраслевой системе мониторинга рыболовства (ОСМ) не вызывает сомнений. Ранее в работе (Василец и др., 2017) было показано, что при визуальной классификации треков судна по данным спутникового позиционирования можно выявить явные противоречия в судовых суточных донесениях. В качестве одного из таких методов контроля данных ССД может быть предложена "видеофиксация" уловов, при этом необходимо отметить, что реальных отечественных систем спутниковой связи для оперативного востребования данной информации в настоящее время не существует.

С другой стороны, в статье (Василец и др., 2018) , был предложены основные этапы метода выявления подозрительной отчетности о вылове на основе сопоставления позиционной и промысловой информации:

- автоматическая классификация трека судна с целью выделить фрагменты, когда судно проводит операции по добыче вылову;
- по координатам определенных на предыдущем этапе фрагментов определяются промысловые зоны;
- далее оценивается время промысловых усилий в каждой из отчетных промысловых зон;
- определяется отношение вылова к времени затраченному на промысел в каждой из анализируемых промысловых зон;
- по отклонению вышеуказанного отношения для судна от среднего для промысловой зоны и сезона добычи (вылова), то делается вывод достоверности отчетности рассматриваемого судна.

Таким образом, внедрение предложенной методики основывается на развитии автоматической классификации треков судна (Дегай и др., 2017) с привязкой периодов траления к промысловым зонам. Возможность автоматической классификации треков судов повысилась благодаря достаточно высокой частоте позиционного опроса судна. Использование терминалов АИС и ГОНЕЦ в составе судового технического средства контроля ОСМ позволило с помощью спутниковой составляющей системы мониторинга ориентировочно в несколько раз (от четырех до семи) увеличить частоту спутникового позиционного опроса судна.

В работе (Пырков и др., 2018) представлены результаты первых опытов автоматизации комплексного анализа позиционной и промысловой информации.

Для выделения периодов траления был выбран простейший подход – считалось, что судно проводит траловые работы, если его скорость больше двух и меньше пяти узлов. Было показано, что даже простейшая автоматизация классификации с последующими действиями вышеуказанной методики:
- по координатам определенных на предыдущем этапе фрагментов определяются промысловые зоны;
- далее оценивается время промысловых усилий в каждой из отчетных промысловых зон;
позволяет выявить серьезные рассогласования между отчетностью судна и треками определенными с помощью технических средств контроля.

Таблица 1

Обобщающий тип	Тип судна
БМРТ	БМРТ несерийн.
БМРТ	БМРТ типа "Иван Бочков" пр.В-408
БМРТ	БМРТ типа "Прометей"
БМРТ	БМРТ типа "Пулковский Меридиан" пр.1288
БМРТ	БМРТ типа "Сотрудничество" пр.Д-1305
РТМ	РТМ несерийн.
СРТМ	СРТМ типа "Иван Шаньков пр.2198"
СРТМ	СРТМ несерийн.
СРТМ	СРТМ типа "Андрей Смирнов" пр.503М
СРТМ	СРТМ типа "Баренцево море" пр.1332
СРТМ	СРТМ типа "Василий Яковенко" пр.5023М
СРТМ	СРТМ типа "Железный поток" пр.502Э
СРТМ	СРТМ типа "Мыс Корсакова" пр.FVS-419
СРТМ	СРТМ типа "Оболонь" пр.333
СРТМ	СРТМ типа "Стеркодер" пр.Р-8830
СРТМ	СРТМ типа "Ягры" пр.50010
МРКТ	МРКТ несерийн.
КРПС	КРПС несерийн.
КРПС	КРПС типа "Виндаис"
МКРТМ	МКРТМ типа "Лаукува" пр.12961
МРТР	МРТР несерийн.
МРТР	МРТР типа "Гирулял" пр.1296
МмРТР	МмРТР типа "Балтика" пр.1328
ТР	ТР несерийн
ТР	ТР пр.12913TR
ТР	ТР типа "50 лет СССР" пр.1347
ТР	ТР типа "Бухта Русская" пр.13476
ТР	ТР типа "Комсомолец Приморья"
ТР	ТР типа "Радужный" пр.1350
НИС	НИС типа "Гидробиолог" пр.3936
НИС	НИС типа "Профессор Марти" пр.833
РПС	РПС типа "Вильнюс" пр.1332
МКРТМ	МКРТМ несерийн.

В работе (Белоконь и др., 2019) показано, что повышение точности автоматической классификации необходимо для проведения корректного сопоставления спутниковых позиционных данных и отчетности пользователей ОСМ.

Следует отметить, что во всех вышеуказанных работах классификация треков проводилась на примере треков, для которых частота определения спутниковых позиций судов приблизительно позиция в пятнадцать минут. Однако, даже в настоящее время для многих судов и районов промысла характерна частота позиционного опроса в два и даже четыре раза ниже.

В данном докладе рассматривается возможность применения методов машинного обучения для классификации треков судов по спутниковым позиционным данным. Оценивается эффективность использования методов машинного обучения для автоматической классификации треков судов, в том числе на участках трека с невысокой частотой позиционного опроса. Для применения методов машинного обучения необходимо подготовить значительный объем данных для обучения и проверки. Для подготовки вышеуказанных данных были проведены предварительные мероприятия и исследования.

В первую очередь были выбраны только такие треки судов которые находятся на значительном расстоянии (не менее пятидесяти километров) от портов.

Далее среди оставшихся были выделены сутки, в которых у испытуемых судов было более пятидесяти позиций. Данное условие обеспечивало высокую вероятность того, что частота позиционного опроса не менее двух позиций в час. В результате были выделены треки для типов судов, которые представлены в Таблице 1. Для первоначальной отработки алгоритма машинного обучения был выбран тип судов – Большой морозильный траулер (БМРТ) проект 1288 "Пулковский Меридиан".

Предварительная попытка классификации треков судов по скорости судна оказалась неудачной. Для выяснения причин проблемы была построена гистограмма скоростей судов по значениям, которые указываются в пакете, получаемом судовым техническим средством контроля, представленная на Рисунке 1.

Рисунок 1

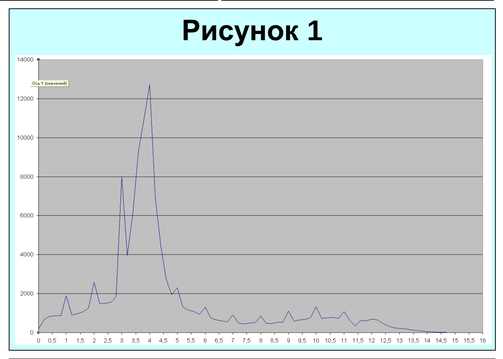
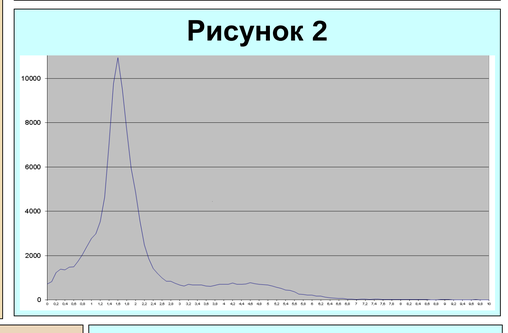


Рисунок 2



Из вида вышеуказанного графика можно сделать вывод, что главной причиной проблем при классификации фрагментов трека судна по скорости являются многочисленные сильные выбросы. Для устранения данной проблемы была построена гистограмма скоростей судна, которые определялись по координатам судна (Рисунок 2). По виду последней гистограммы были выделены интервалы скоростей, по которым далее проводилась предварительная классификация треков судна, представленные в Таблице 2.

Таблица 2

индекс	интервал скоростей	цвет
0	- менее 0,8 узлов	синий
1	- больше 0,8 и меньше 1,5 узлов	желтый
2	- больше 1,5 и меньше 2,5 узлов	коричневый
3	- больше 2,8 и меньше 5,5 узлов	зеленый
4	- больше 5,5 узлов	розовый

Для большей определенности предварительной классификации треков исследовались перемещения судов которые проводили добычу-вылов в Охотском море. Среди всего множества позиций судов типа БМРТ "Пулковский Меридиан» выбирались такие фрагменты треков, когда у судна больше пяти дней подряд в сутки позиция определялась более пятидесяти раз. В рамках выделенных тестовых треков, используя результаты предварительной классификации, визуально определялись участки, когда судно осуществляло траление.

Для удобства визуального выделения участков траления на диаграмме, вспомогательная программа выводит от семи до пятнадцати последовательных фрагментов трека относящихся к разным классам Таблицы 2.

На Рисунке 3 и Рисунке 4 представлены примеры диаграмм, по которым визуально выделялись фрагменты траления. На Рисунке 3 представлены семь последовательных фрагментов трека относящихся к индексам классификации 1 и 2. Все семь фрагментов отнесены к тралению. На Рисунке 4 представлены пятнадцать последовательных фрагментов трека относящихся к индексам классификации 1, 2 и 3. Пять фрагментов сверху слева отнесены к переходу. Десять предшествовавших фрагментов (справа внизу) отнесены к тралению.

Таким образом, в результате визуального выделения фрагментов траления была получена обучающая выборка содержащая положительных примеров: 10462, отрицательных 6050.

Были проведены тесты на основе модели классификации базового уровня (baseline model) с помощью алгоритмов случайный лес и бустинг (реализация - библиотека xgb). Оптимизируется метрика accuracy (Accuracy = (TP+TN)/(TP+TN+FP+FN) where: TP = True positive; FP = False positive; TN = True negative; FN = False negative).

Для борьбы с переобучением используем метод перекрёстной проверки (cross validation) с разбиением на 3 части.

В зависимости от алгоритма и его параметров результат 70% - 65%. Что позволяет говорить о хорошем уровне подготовки обучающей выборки.

Рисунок 3

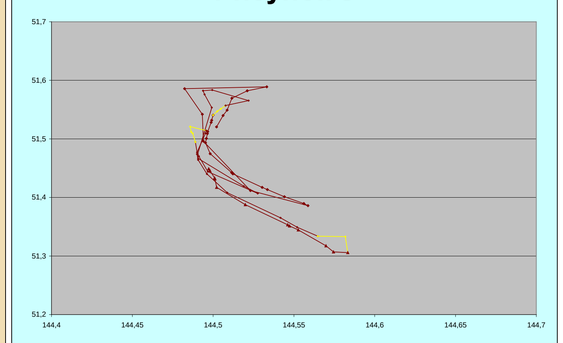
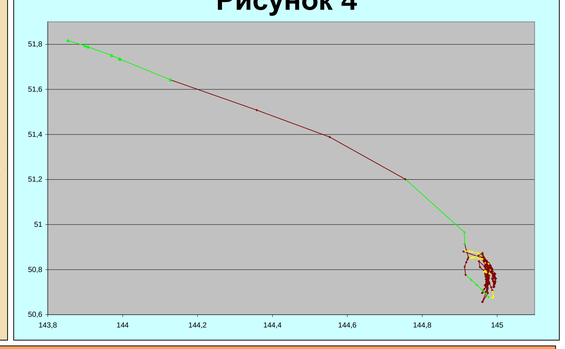


Рисунок 4



Выводы

Разработан алгоритм подготовки обучающей выборки. Разработаны методики проведения обучения с целью последующего выявления фрагментов треков судов когда осуществляется траление. На основе вышеуказанных разработанных методик в ближайшее время будут проведены работы по расширению обучающей выборки, а также дальнейшего совершенствования методики подготовки набора данных и методики проведения машинного обучения.

Работы проводились с использованием технологий и данных центра коллективного пользования ЦКП "ИКИ-Мониторинг" (Лупян и др., 2015).
Работа выполнена при поддержке ФАНО (тема "Мониторинг", госрегистрация № 01.20.0.2.00164).

Литература

1. *Василец П.М., Терентьев Д.А., Пырков В.Н.* Анализ пространственного распределения вылова камбал за 2011 - 2016 годы в прикамчатских и смежных с ними районах по данным отраслевой системы мониторинга Росрыболовства // Исслед. водн. биол. ресурсов Камчатки и сев.-зап. части Тихого океана: Сб. науч. тр. Камчат. НИИ рыб. хоз-ва и океанографии. 2017. Вып. 47. С. 65-76. DOI: 10.15853/2072-8212.2017.47.65-76.
2. *Василец П.М., Терентьев Д.А., Коробов С.А., Пырков В.Н., Солодилов. А.В., Дегай А.Ю.* Возможности оценки достоверности пространственной информации по вылову донных рыб в отраслевой системе мониторинга Росрыболовства на основе комплексного анализа производственно-промысловой отчетности и данных спутникового позиционирования промысловых судов // Современные проблемы дистанционного зондирования Земли из космоса. 2018. Т. 15. № 7. С. 35-42. DOI: 10.21046/2070-7401-2018-15-7-35-42.
3. *Дегай А.Ю., Пырков В.Н., Черных В.Н., Солодилов А.В.* Рассмотрение возможности классификации треков в системе мониторинга рыболовства с использованием современных технических средств контроля // Пятнадцатая всероссийская открытая конференция "Современные проблемы дистанционного зондирования Земли из космоса". Москва. ИКИ РАН, 13-17 ноября 2017.Тезисы докладов, 2017. С. 82.
4. *Пырков В.Н., Василец П.М., Дегай А. Ю, Андреев М.В., Черных В.Н., Солодилов А.В.* Новый подход к верификации отчетов о вылове на основе классификации треков спутникового позиционирования в системе мониторинга рыболовства // Информационные технологии в дистанционном зондировании Земли - RORSE 2018. ИКИ РАН, 2019. С. 145-149. DOI: 10.21046/rorse2018.145.
5. *Белоконь З.С., Василец П.М., Дегай А.Ю. , Пырков В.Н.* Возможность повышения точности автоматической классификации треков судов в Отраслевой системе мониторинга Росрыболовства с помощью методики выявления периодов сближения судов// XVI всероссийская научно-техническая конференция "Современные методы и средства океанологических исследований" (МСОИ-2019). М.:ИД Академии Жуковского, 2019. Т.I. С. 251-254.
6. *Лупян Е.А., Прошин А.А., Бурцев М.А., Балашов И.В., Барталев С.А., Ефремов В.Ю., Кашницкий А.В., Мазуров А.А., Матвеев А.М., Суднева О.А., Сычугов И.Г., Толпин В.А., Уваров И.А.* Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 263-284.

Институт Космических Исследований РАН 117997, Москва, Профсоюзная ул., д. 84/32
тел. +7-495-3331077, факс +7-495-9133040 E-mail: info@smis.iki.rssi.ru