

# TS-CP2 for semi-supervised change point detection

Андрей Веприков    Алексей Зайцев    Евгения Романенкова    Александр Степкин

Лето 2022

# Содержание

- 1 Методы детектирования и способ их совмещения
- 2 Датасеты
- 3 Эксперименты
- 4 Finetuning

# Введение

Разладка во временном ряду – это момент резкой смены вероятностного распределения составляющих его данных. Интересной задачей машинного обучения и нейронных сетей является детектирование этих самых разладок в различных последовательностях.

Этим летом я совместил 2 метода детектирования разладок в одну модель, исследовал качество этой модели на различных датасетах (Syntetic 1D и MNIST) и выбрал оптимальные гиперпараметры для этой модели.

Какие именно методы я совместил и каким образом, особенности датасетов, об этом всем будет подробнее рассказано далее.

## Методы детектирования и способ их совмещения

# Unsupervised метод TS-CP2

В основе этого подхода лежит кодировщик (encoder), который отображает «позитивные» пары (то есть пары из одного распределения) в метрическое пространство так, чтобы между этими парами расстояние было небольшим, а «негативные» пары (то есть пары из разного распределения) он отображает в элементы, расстояние между которыми достаточно большое.

Ключевым является семплирование положительных и негативных пар. В этом методе позитивными считаются пары, которые взяты из близких мест в последовательности, а негативными – пары из различных последовательностей в батче.

# Supervised метод TS-CP2

Этот метод похож на предыдущий, разница только в семплировании положительных и негативных пар – здесь мы уже пользуемся разметкой.

# Комбинирование этих методов

При обучении нашей модели на каждом тренировочном шаге мы брали батчи из двух разных датасетов – размеченного и неразмеченного, параллельно вычислялись функции ошибок, назовем их `unsupervised loss` и `supervised loss`. Финальная функция ошибки вычислялась как  $\alpha * \text{unsupervised loss} + (1-\alpha) * \text{supervised loss}$ . Гиперпараметр  $\alpha$  лежит в промежутке  $[0; 1]$  и регулирует вклад каждого метода в нашу модель.

# Датасеты



# Syntetic 1D

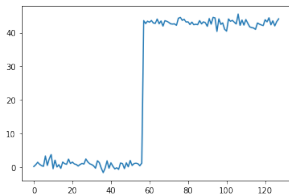
В качестве простейшей выборки были сгенерированы последовательности длины 128, состоящие из 1-мерных гауссовских случайных величин.

Они имеют вид

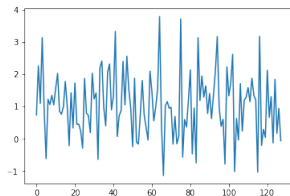
$$x_1, \dots, x_\theta \sim N(0, 1) \quad x_{\theta+1}, \dots, x_{128} \sim N(\mu, 1),$$

где  $\mu \in \{1, \dots, 100\}$  – случайное математическое ожидание, а  $\theta \in \{2, \dots, 127, 128\}$  – случайный момент разладки (если  $\theta = 128$ , значит, разладки нет).

Примеры последовательностей из Syntetic 1D:



С разладкой



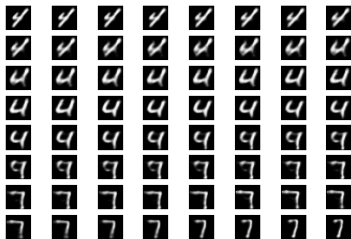
Без разладки

# MNIST

Данная выборка была получена на основе базы данных MNIST, которая представляет собой набор изображений рукописных арабских цифр размера  $28 \times 28$  пикселей.

Если есть разладка, то наши последовательности состоят только из изображений цифры 4, а если она есть, то в какой-то случайный момент 4 переходят в 7.

Примеры последовательностей из MNIST:



С разладкой



Без разладки

# Эксперименты

# Варьирование $\alpha$ и supervised num

В первом эксперименте я одновременно менял  $\alpha$  и supervised num – размер размеченного датасета, для неразмеченного датасета размер был фиксирован – 500 последовательностей.

Для каждого значения двух гиперпараметров я обучал модель 3 раза при разных седах и усреднял результаты.

Ниже представлены значения для метрики F1-Score на двух датасетах.

# Варьирование $\alpha$ и supervised num, таблицы

	$s = 5$	$s = 10$	$s = 15$	$s = 25$	$s = 50$	$s = 75$
$\alpha = 0$	$0.8950 \pm 0.0108$	$0.9056 \pm 0.0225$	$0.9239 \pm 0.0091$	$0.9369 \pm 0.0192$	$0.9406 \pm 0.0142$	$0.9322 \pm 0.0219$
$\alpha = 0.3$	$0.8642 \pm 0.0404$	$0.9087 \pm 0.0097$	$0.9045 \pm 0.0100$	$0.9154 \pm 0.0117$	$0.9245 \pm 0.0017$	$0.9183 \pm 0.0084$
$\alpha = 0.5$	$0.8914 \pm 0.0193$	$0.8988 \pm 0.0125$	$0.8900 \pm 0.0118$	<b><math>0.9198 \pm 0.0032</math></b>	$0.8525 \pm 0.0420$	$0.8775 \pm 0.0340$
$\alpha = 0.7$	$0.7989 \pm 0.1254$	$0.8882 \pm 0.0423$	$0.7502 \pm 0.1500$	$0.7602 \pm 0.0953$	$0.7336 \pm 0.0276$	$0.7831 \pm 0.0626$

Syntetic 1D

	$s = 15$	$s = 20$	$s = 25$	$s = 30$
$\alpha = 0$	$0.8620 \pm 0.0482$	$0.8610 \pm 0.0699$	$0.9538 \pm 0.0119$	$0.9612 \pm 0.0193$
$\alpha = 0.3$	$0.6758 \pm 0.0882$	$0.7126 \pm 0.0102$	$0.7123 \pm 0.0584$	$0.6801 \pm 0.0340$
$\alpha = 0.5$	$0.6495 \pm 0.0408$	$0.7152 \pm 0.0279$	$0.7253 \pm 0.0195$	<b><math>0.7193 \pm 0.0301</math></b>
$\alpha = 0.7$	$0.6845 \pm 0.0260$	$0.6665 \pm 0.0278$	$0.7043 \pm 0.0079$	$0.7107 \pm 0.0158$

MNIST

# Варьирование unsupervised num

Во втором эксперименте я менял unsupervised num – размер не размеченного датасета, alpha и supervised num были выбраны исходя из предыдущего эксперимента (эти значения помечены жирным в таблицах).

Ниже представлены значения для метрики F1-Score на двух датасетах.

unsupervised_num	F1-Score
400	0.9147 ± 0.0033
450	0.8724 ± 0.0463
<b>500</b>	<b>0.9198 ± 0.0032</b>
550	0.7942 ± 0.1137
600	0.8804 ± 0.0431

Syntetic 1D

unsupervised_num	F1-Score
400	0.6814 ± 0.0299
450	0.6540 ± 0.0427
500	0.7193 ± 0.0301
<b>550</b>	<b>0.7436 ± 0.0261</b>
600	0.7091 ± 0.0111

MNIST

# Построение графиков обучения

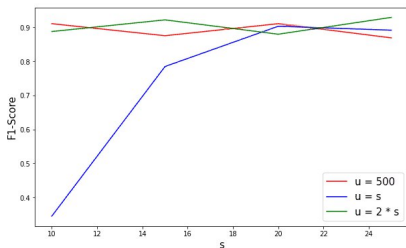
Третьим экспериментом было построить график зависимости F1-Score для различных supervised num, причем unsupervised num выбиралось в первом случае исходя из предыдущего эксперимента, во втором оно бралось равным supervised num, а в третьем  $2 * \text{supervised num}$ . Значения alpha выбирались исходя из предыдущих экспериментов.

Ниже представлены таблицы и графики, которые из них получились

# Построение графиков обучения, таблицы и графики

	$s = 10$	$s = 15$	$s = 20$	$s = 25$
$u = 500$	0.9103	0.875	0.9103	0.8687
$u = s$	0.3455	0.7846	0.9028	0.8912
$u = 2 * s$	0.8873	0.9215	0.8794	0.9288

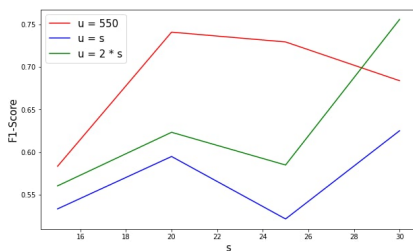
Syntetic 1D



Syntetic 1D

	$s = 15$	$s = 20$	$s = 25$	$s = 30$
$u = 550$	0.5833	0.7407	0.7292	0.6837
$u = s$	0.5333	0.5948	0.5215	0.625
$u = 2 * s$	0.5603	0.6232	0.5848	0.7556

MNIST



MNIST



# Варьирование num change

Отдельно для датасета MNIST я менял num change в неразмеченном датасете – кол-во последовательностей с разладкой, без разладки было 550 - num change последовательностей. supervised num и alpha были фиксированы – 30 (15 с разладкой, 15 без) и 0.5 соответственно.

Ниже представлена таблица, которая у меня получилась

num_change	F1-Score
100	0.7518 $\pm$ 0.0191
150	0.7217 $\pm$ 0.0193
200	0.6584 $\pm$ 0.0238
225	0.7193 $\pm$ 0.0301
300	0.7106 $\pm$ 0.0300
350	0.7811 $\pm$ 0.0239
400	0.6473 $\pm$ 0.0774
450	0.6738 $\pm$ 0.0449

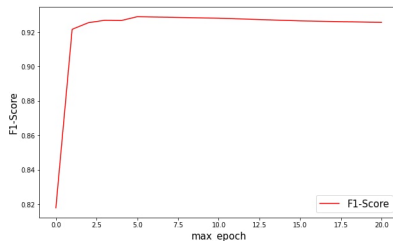
# Finetuning

# Описание

Теперь отойдем от идеи параллельного обучения и будем использовать Finetuning – мы сначала обучим ее только на неразмеченных данных (unsupervised num = 500), а потом уже дообучим ее на размеченных данных. Изначально мы обучали модель на неограниченном кол-ве эпох – пока процесс не остановится сам, а уже дообучали на конкретном числе эпох, потом строили график зависимости F1-Score от кол-ва эпох. Также мы меняли supervised num – размер дообучающей выборки. Ниже представлены таблицы и графики, которые у нас получились

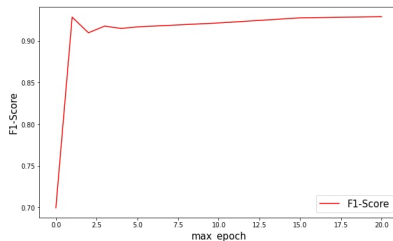
# Syntetic 1D

max_epoch	AUC	F1-Score	COVER
0	1350.2109 ± 0.0000	<b>0.8178 ± 0.0000</b>	0.9313 ± 0.0000
1	1134.0923 ± 40.5511	<b>0.9215 ± 0.0030</b>	0.9739 ± 0.0002
2	1041.0715 ± 19.6686	<b>0.9253 ± 0.0002</b>	0.9750 ± 0.0004
3	1015.4839 ± 21.7968	<b>0.9267 ± 0.0015</b>	0.9763 ± 0.0009
4	990.0316 ± 1.0131	<b>0.9266 ± 0.0016</b>	0.9764 ± 0.0005
5	982.5919 ± 7.2966	<b>0.9288 ± 0.0000</b>	0.9772 ± 0.0002
10	991.7048 ± 16.1740	<b>0.9279 ± 0.0012</b>	0.9769 ± 0.0011
15	986.8419 ± 4.0656	<b>0.9264 ± 0.0017</b>	0.9766 ± 0.0006
20	1018.8196 ± 13.0035	<b>0.9255 ± 0.0002</b>	0.9769 ± 0.0003



supervised num = 500

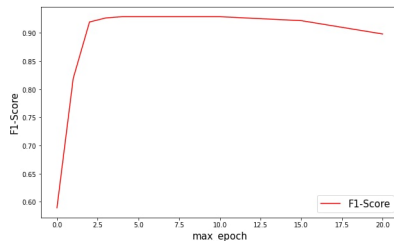
max_epoch	AUC	F1-Score	COVER
0	1465.7275 ± 0.0000	0.6997 ± 0.0000	0.8970 ± 0.0000
1	935.7097 ± 91.7876	0.9285 ± 0.0293	0.9258 ± 0.0351
2	1088.0619 ± 19.9490	0.9097 ± 0.0022	0.8914 ± 0.0026
3	1111.2375 ± 1.4985	0.9177 ± 0.0017	0.8937 ± 0.0006
4	1115.3999 ± 8.8529	0.9149 ± 0.0005	0.8943 ± 0.0005
5	1105.1944 ± 8.8972	0.9168 ± 0.0034	0.8941 ± 0.0011
10	1098.1161 ± 9.3388	0.9215 ± 0.0052	0.8956 ± 0.0022
15	1095.1049 ± 6.4511	0.9276 ± 0.0017	0.8971 ± 0.0014
20	1094.2226 ± 5.9605	0.9290 ± 0.0028	0.8981 ± 0.0017



supervised num = 250

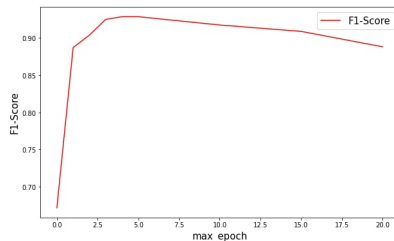
# Syntetic 1D

max_epoch	AUC	F1-Score	COVER
0	1481.9738 $\pm$ 0.0000	<b>0.5893 <math>\pm</math> 0.0000</b>	0.8886 $\pm$ 0.0000
1	1274.5978 $\pm$ 66.6809	<b>0.8193 <math>\pm</math> 0.1038</b>	0.9438 $\pm$ 0.0321
2	1197.7415 $\pm$ 8.8820	<b>0.9194 <math>\pm</math> 0.0015</b>	0.9772 $\pm$ 0.0008
3	1145.2041 $\pm$ 26.6605	<b>0.9266 <math>\pm</math> 0.0032</b>	0.9793 $\pm$ 0.0006
4	1038.3662 $\pm$ 14.1210	<b>0.9288 <math>\pm</math> 0.0000</b>	0.9802 $\pm$ 0.0001
5	1046.2301 $\pm$ 6.6570	<b>0.9288 <math>\pm</math> 0.0000</b>	0.9803 $\pm$ 0.0000
10	1067.5729 $\pm$ 19.6342	<b>0.9288 <math>\pm</math> 0.0000</b>	0.9801 $\pm$ 0.0002
15	1099.2345 $\pm$ 8.3709	<b>0.9218 <math>\pm</math> 0.0030</b>	0.9772 $\pm$ 0.0010
20	1154.5492 $\pm$ 12.5719	<b>0.8981 <math>\pm</math> 0.0034</b>	0.9682 $\pm$ 0.0016



supervised num = 100

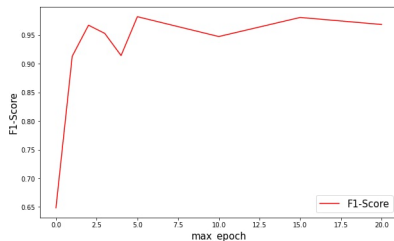
max_epoch	AUC	F1-Score	COVER
0	1490.0293 $\pm$ 0.0000	<b>0.6716 <math>\pm</math> 0.0000</b>	0.8923 $\pm$ 0.0000
1	1240.3684 $\pm$ 34.5952	<b>0.8873 <math>\pm</math> 0.0207</b>	0.9615 $\pm$ 0.0090
2	1199.6192 $\pm$ 4.5594	<b>0.9040 <math>\pm</math> 0.0036</b>	0.9709 $\pm$ 0.0018
3	1189.3978 $\pm$ 9.0754	<b>0.9252 <math>\pm</math> 0.0030</b>	0.9781 $\pm$ 0.0008
4	1127.9910 $\pm$ 13.3004	<b>0.9288 <math>\pm</math> 0.0000</b>	0.9802 $\pm$ 0.0001
5	1131.0608 $\pm$ 15.8516	<b>0.9288 <math>\pm</math> 0.0000</b>	0.9804 $\pm$ 0.0000
10	1163.7966 $\pm$ 3.6990	<b>0.9177 <math>\pm</math> 0.0092</b>	0.9755 $\pm$ 0.0029
15	1132.4999 $\pm$ 19.2155	<b>0.9090 <math>\pm</math> 0.0070</b>	0.9728 $\pm$ 0.0025
20	1048.8060 $\pm$ 32.0784	<b>0.8883 <math>\pm</math> 0.0135</b>	0.9673 $\pm$ 0.0048



supervised num = 50

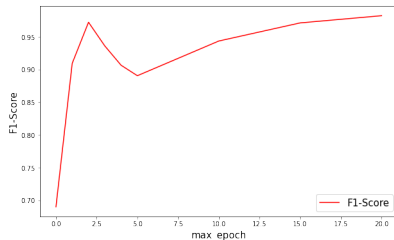
## MNIST

max_epoch	AUC	F1-Score	COVER
0	326.7924 ± 0.0000	<b>0.6485 ± 0.0000</b>	0.8110 ± 0.0000
1	206.4070 ± 50.0186	<b>0.9129 ± 0.0065</b>	0.8833 ± 0.0129
2	89.8429 ± 16.0769	<b>0.9671 ± 0.0219</b>	0.9081 ± 0.0095
3	104.2689 ± 11.8005	<b>0.9527 ± 0.0313</b>	0.9194 ± 0.0032
4	201.449 ± 60.5243	<b>0.9141 ± 0.0667</b>	0.8974 ± 0.0086
5	76.1339 ± 2.4245	<b>0.9819 ± 0.0137</b>	0.9194 ± 0.0053
10	111.9123 ± 48.7905	<b>0.9473 ± 0.0302</b>	0.9215 ± 0.0014
15	183.3653 ± 63.9131	<b>0.9806 ± 0.0177</b>	0.9198 ± 0.0036
20	74.5935 ± 2.5516	<b>0.9683 ± 0.0399</b>	0.9196 ± 0.0022



supervised num = 500

max_epoch	AUC	F1-Score	COVER
0	300.7945 ± 0.0000	<b>0.6899 ± 0.0000</b>	0.8213 ± 0.0000
1	219.6781 ± 23.7464	<b>0.9093 ± 0.0074</b>	0.8517 ± 0.0103
2	150.6444 ± 33.3137	<b>0.9720 ± 0.0153</b>	0.9044 ± 0.0098
3	82.7836 ± 5.8121	<b>0.9358 ± 0.0285</b>	0.9196 ± 0.0004
4	93.2684 ± 20.9937	<b>0.9063 ± 0.0595</b>	0.9188 ± 0.0002
5	106.9713 ± 27.4376	<b>0.8903 ± 0.0265</b>	0.9209 ± 0.0010
10	180.3521 ± 39.3568	<b>0.9433 ± 0.0312</b>	0.9210 ± 0.0014
15	194.214 ± 72.4240	<b>0.9710 ± 0.0079</b>	0.9008 ± 0.0290
20	122.3874 ± 20.0243	<b>0.9820 ± 0.0108</b>	0.9195 ± 0.0004



supervised num = 250