Aalto University
School of Science
Bachelor's Programme in Science and Technology

Video change point detection based on viewing statistics

Bachelor's Thesis

April 17, 2022

Saskia Kivistö

Aalto University
School of Science
Bachelor's Programme in Science and Technology

ABSTRACT OF
BACHELOR'S THESIS

Author:	Saskia Kivistö
Title of thesis:	Utilisation of viewing statistics in detecting NPVR closing credits
Date:	April 17, 2022
Pages:	
Major:	Computer Science
Code:	SCI3027
Supervisor:	Assistant Professor Parinya Chalermsook
Instructor:	Wanchote Jiamjitrak, M.Sc. (Tech) (Department of Computer Science)
Abstract	
Keywords:	change point detection,
Language:	English

Aalto-yliopisto
Perustieteiden korkeakoulu
Teknistieteellinen kandidaattiohjelma

KANDIDAATINTYÖN TIIVISTELMÄ

Tekijä:	Saskia Kivistö
Työn nimi:	Lopputekstien paikannus NPVR tallenteesta katsomistilastoja hyödyntäen
Päiväys:	17. huhtikuuta 2022
Sivumäärä:	
Pääaine: **Koodi:**	Tietotekniikka SCI3027
Vastuuopettaja: **Työn ohjaaja(t):**	Apulaisprofessori Parinya Chalermsook DI Wanchote Jiamjitrak (Tietotekniikan laitos)
Tiivistelmä	
Avainsanat:	muutospisteiden havaitseminen,
Kieli:	englanti

Contents

1 Introduction

Watching television used to be a very time-sensitive activity. If one wanted to see a TV program, they had to watch it when it was broadcasted. Videocassette recorder brought the freedom to store programs and to choose freely when to watch them. Even though newer technologies have replaced the videocassette recorder, one major inconvenience still remains. Programs are not broadcasted strictly according to the schedule of the Electronic program guide (EPG). Sometimes programs start earlier or end later than what is stated in the EPG. To ensure that the entire program is recorded, recording must start before the EPG start time and end after the EPG end time. This usually results in some non-program content being included in the recording. Skipping over the non-program content can be frustrating for the person watching the recording.

Knowing when the program truly starts and ends would solve the issue, but this information is not generally available. This leaves the option to detect the start and end of the program from the recording. Various artificial intelligence solutions have been devised to solve the problem, for example [2] [3] [5] [9], but as program content can be quite varied it is difficult to find a universal solution.

Network personal video recorder (NPVR) is a type of service for recording broadcast TV programs for later viewing. Instead of storing recordings on the users' local device, NPVR stores recordings on the content provider's server. The start and end times of recordings are determined by the scheduling information given in the EPG, and a fixed amount of margin is added on both ends to ensure that the entire program is recorded. This leads to non-program content being included in recordings, which the users typically skip. However, statistics of which parts of the recording users watch and which parts they skip can be collected and analysed.

The goal of this thesis is to study whether user viewing behaviour can be used to detect when the actual program content begins and ends in an NPVR recording. I am writing this thesis for an NPVR service provider company. From the perspective of an NPVR service provider, detecting the location of core program content is useful for the following reasons. Firstly, less storage space is needed if the surplus content is discarded. Secondly, it is convenient for the customers if the relevant content of a recording is pre-identified and they do not have to search for it.

The thesis is structured as follows. Section 2 gives an overview on the characteristics of the viewing behavioiur data. Section 3 discusses the theoretical background of signal change point detection from the perspective of this specific use case. Section 4 discusses the Python scientific library `ruptures` and how it is used to detect change points in this thesis. The results are evaluated in section 5. Section 6 considers the viability of using user viewing behaviour for change point detection, based on the previous sections. Lastly, the main points of this thesis are summarised in section 7.

2 User viewing behaviour data

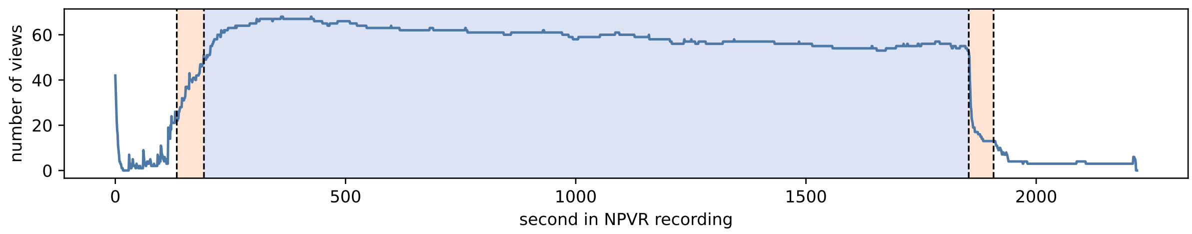
Whenever an NVPR video is watched by a user, certain metrics about the viewing event are saved. The main reason for collecting viewing metrics is the monitoring of the amount of views and the user experience quality. It can be calculated from the metrics of one view which parts of the video the user watched and which parts they skipped. Aggregating this data from multiple views of the same recording allows acquiring an overview of what parts of the recording users typically watch. This type of recording specific aggregated view count is referred to as user behaviour data in this thesis.

The content of an NVPR recording can be divided into program and non-program content. Non-program content is the surplus material at the very beginning and end of a recording, that is a byproduct of ensuring that the entire program is recorded even if its broadcast time deviates sligtly from the EPG schedule. Program content can be categorised into opening and closing credits, core program content and advertisement breaks, although not all programs have them.

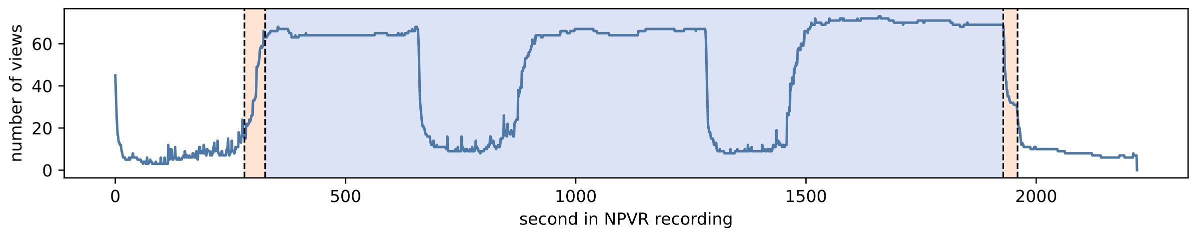
User interest for the content categories varies. Core program content is what the users watch the recordings for, and non-program content is irrelevant for the users. This is reflected in the viewing behaviour. Users also tend to skip over advertisement breaks. Viewing behaviour regarding the opening and closing credits is not as clear cut, but generally users are content with starting watching during the opening creidts. Closing credits do not typically interest users, given there is no extra content included in them, such as a preview of the next episode.

Two examples of user viewing behaviour data are illustrated in Figure 1. The data in both figures consists of a sample of 100 user views, but the views are from two different TV programs. Figure 1a views are from a sitcom episode and Figure 1b views are from an episode of a soap opera. The episodes are divided into one second-long segments, and calculated for each segment is the number of user views in which the segment was watched. The segments are plotted on the horizontal axis, and the number of user views per segment is plotted on the vertical axis. For example, 63 users from the sample of 100 users watched the part of the sitcom episode between 0:10:00 - 0:10:01 (600 on the horisontal axis in Figure 1a).

The background colour in the figure represents the content type of the time segment: non-program content is white, core program content is blue and beige indicates the opening and closing credits. The change points are marked with a vertical dashed line. There is a distict increase in the number of views during the opening credits in both figures.



(a) 30 min sitcom episode without advertisements, 100 views



(b) 30 min soap opera episode with two advertisement breaks, 100 views

Figure 1: Visualisation of user viewing behaviour for two example recordings

Likewise, there is a decrease in the number of views during closing credits. The program visualised in Figure 1b has two advertisement breaks, which appear as recesses in the number of views during the core program content. Figure 1a episode has no advertisement breaks.

2.1 Data cleaning

This section explains how the sample views are chosen. An uniform sample size is used for all recordings in order to simplify result evaluation and comparison. The characteristic pattern for viwing behaviour described in the end of the previous section is usually visible from a dozen views. Sample size of 100 views was chosen because it should be large enough to ensure that it is very unlikely that the characteristic pattern does not emerge.

As the first step, all views of a recording are sorted from oldest to newest. The views will be added to the sample in this order, but potentially incorrect or irrelevant views are filterd out. This means that the sample views are not simply the 100 earliest ones. Firstly, views which end before the recording process ends are discarded. Also views with no skips are discarded, since the change point detection relies on inspecting which parts the viewers skip, so views without skips are useless. Lastly views with implausible player source duration are removed. This includes the cases where player source duration does not exist, it is negative or close to maximum value for unsigned 32-bit integer.

3 Theoretical background

3.1 Signal change detection

Locating video content changes from user viewing behaviour time series data can be formulated as a signal processing problem, more precisely as a change point detection problem. Signal change point detection is quite widely researched topic, since it has applications in multiple fileds such as network traffic data analysis [6] [8], bio-informatics [7] [14] and climatology [10] [13].

Change point detection problems can be divided into two main categories, depending on whether the change detection must be done for incoming data in near real-time, or not. Methods solving the former case are referred to as online algorithms. Offline algorithms solve the latter case, and they differ from the online algorithms by getting the entire dataset as input and typically being more computationally complex, but also by detecting the changes more accurately.

Offline methods can be divided into two categories, based on whether the number of changes in the dataset is known beforehand, or not. If the number of change points is not known, an extra step is needed to determine it. There are multiple methods for doing this.

Literature review by Truong et al. [11] compared different offline change point detection methods. It classified change point detection methods according to how the homogenuity of a signal is measured, and how the segments where to evaluate the homogenuity are chosen from the signal. The measure of homogenuity is referred to as *cost function* and the way to choose the segments is referred to as *search method*.

The literature review divided search methods into two categories, by whether they provide always an optimal solution for change point location with the chosen cost function, or not. It introduced only two optimal search methods, `Opt` and `Pelt`. The difference between them is that `Opt` works only when the number of change points is know beforehand, and `Pelt` is used when the number of change points is unknown.

Different cost functions work well with different types of signals. A cost function that works well for mean shift detection is least squared deviation c_{L_2}. It measures the variance and is defined as follows:

$$c_{L_2}(y_{a.b}) = \sum_{t=a+1}^{b} \|y_t - \overline{y}_{a.b}\|_2^2 \tag{1}$$

3.2 Result evaluation

Standard statistical methods for univariate analysis will be used for result evaluation. The statistics will include five-number summary first devised by Tukey et al. [12], which includes the median, the 1st and 3th quartiles and the minimum and maximum.

4 Methods

4.1 Sample data and ground truth

In order to evaluate how well a method detects change points, a ground truth is needed for comparison. Ground truth can be obtained by having a person look at a video recording and having them write down the timestamps of the change points.

I have collected the start and end times of the change points from 228 NPVR recordings by hand with a margin of error of ± 1 seconds. The sample consists of episodes from 14 different TV series. 138 of the sample recordings were broadcasted on non-commercial channels, and thus do not contain advertisement breaks. Some general information about these recordings is listed in Table 1. The rest 90 recordings with advertisement breaks are listed in Table 2. Every recording in the sample has at least one hundred user views.

# series	# episodes	episode length	genre
1.	6	30 min	reality television
2.	10	30 min	comedy
3.	32	30 min	comedy
4.	9	40 min	drama
5.	4	45 min	drama
6.	16	45 min	drama
7.	30	45 min	drama
8.	8	50 min	drama
9.	20	50 min	drama
10.	3	90 min	drama

Table 1: Sample recordings without advertisement breaks categorized by series

4.2 Change point detection with `ruptures` library

The `ruptures` Python library was used for the change point detection. `ruptures` is based on the findings of a literature reviw conducted by Truong et al. [11] which examined

# series	# episodes	# ad breaks	episode length	genre
11.	35	1	30 min	soap opera
12.	31	2	30 min	soap opera
13.	15	1, 2 or 3	30-60 min	reality television
14.	9	4	90 min	reality television

Table 2: Sample recordings with advertisement breaks categorized by series

various methods for offline change point detection. Selected algorithms examined in the literature review are implemented in `ruptures`.

Choosing the most suitable algorithm for this use-case can be done by considering the three aspects of change detection methods discussed in the literature review: cost-function, search method and constraint. Accuracy is more important than low computational complexity, so optimal method is preferable to an approximate one. `ruptures` has two optimal search methods `Opt` and `Pelt`. `Opt` was first introduced by Bellman [1] for an unrelated problem and `Pelt` was first indroduced by Killick et al. [4].

The `ruptures` library supports three different cost functions that can be used with `Opt` and `Pelt`, c_{L1}, c_{L2} and c_{rbf}.

Following the convention used by Truong et al. the actual change points will be denoted by t and the change points produced by `ruptures` will be denoted by $\hat{t}$ [11]. Each change point has a subscript, where the first letter denotes the credits where the change point belongs to (o: opening credits, c: closing credits). If there are separate change points for the start and end timestamps of the credits, the subscript will have a second letter that denotes which one of those the change point represents (s: start, e: end). This notation is illustrated in Figure 2.

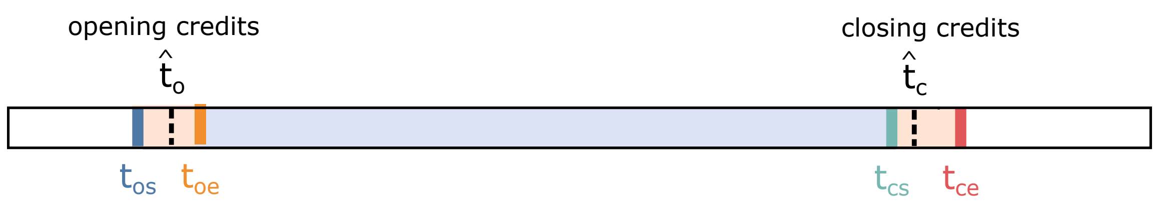


Figure 2: Notation for the actual (t) and predicted ($\hat{t}$) change points

5 Results

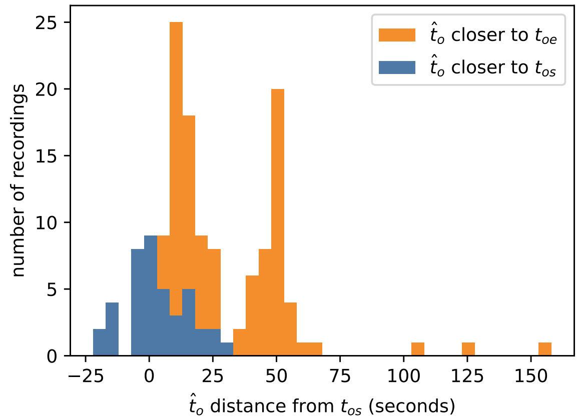
5.1 Non-commercial channels with `Opt` as search method

Change points were calculated with `ruptures` for the 116 recordings without advertisement breaks listed in Table 1. The search method used was `Opt` with a fixed number of change points set to $k = 2$. The cost function used was c_{L_2} as given in equation 1. Statistics of the result are listed in Table 3.

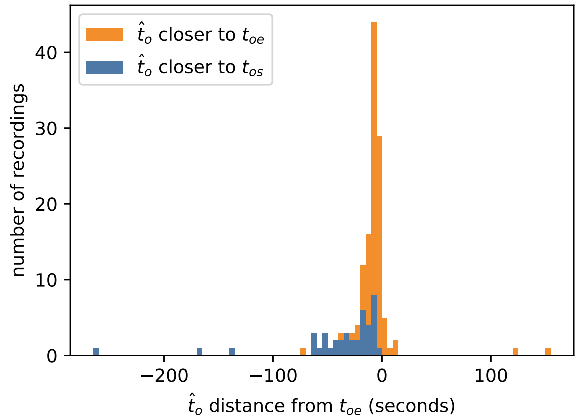
statistic	$\hat{t}_o - t_{os}$	$\hat{t}_o - t_{oe}$	$\hat{t}_c - t_{cs}$	$\hat{t}_o - t_{os}$
minimum	-22	-265	3	-107
1st quartile	9	-17.75	5	-50.75
median	16	-5	7	-35.5
3rd quartile	45,75	-5	8	-28
maximum	157	152	36	5
variance	676	1380	17.2	873
standard deviation	26.0	36.2	4.15	29.5

Table 3: Five-number summary and other statistics for `Opt` c_{L2}, $k = 2$

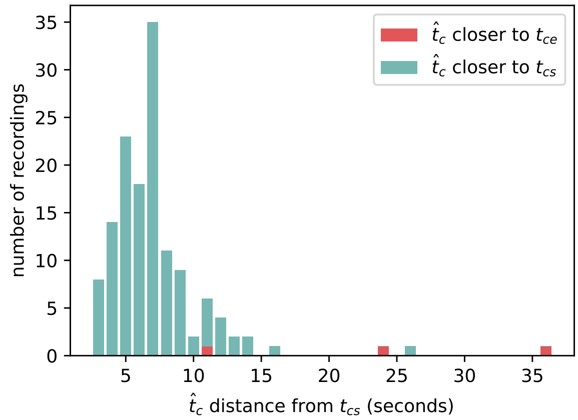
Both `Opt` and `Pelt` are optimal search methods, meaning that with the same cost function the output produced by the methods is identical, if the number of change points predicted by `Pelt` happens to be the same that was given as a parameter k to `Opt`. For example, the `Opt` c_{L_2} output with $k = 2$ for Figure 1a data is visualised in Figure 4a.



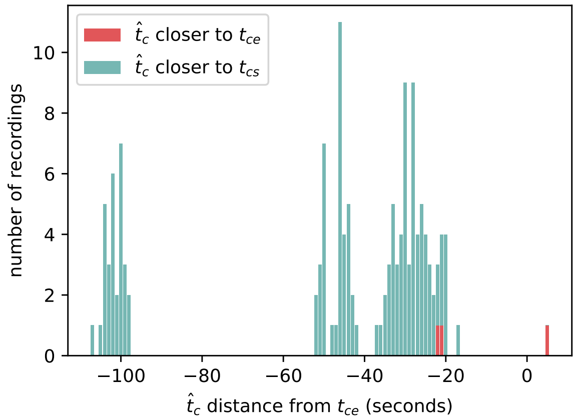
(a) Start of opening credits



(b) End of opening credits



(c) Start of closing credits

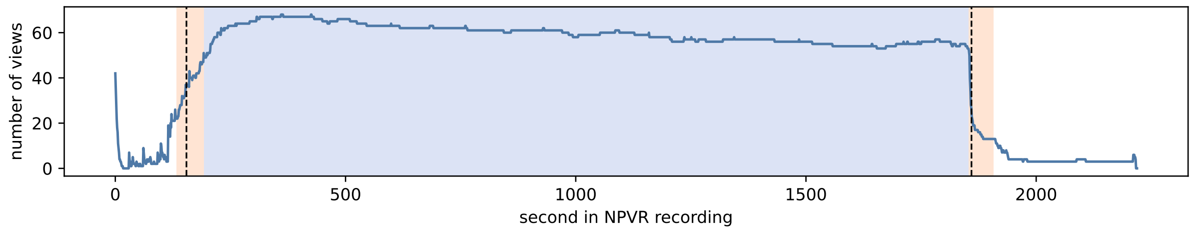


(d) End of closing credits

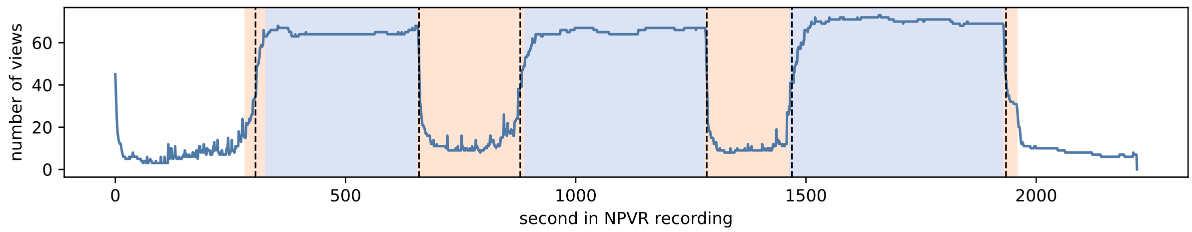
Figure 3: `Opt` c_{L2} output compared to actual change points of credits

5.2 All channels with `Pelt` as search method

An example of `Pelt` output with cost function c_{L2} is visualised in Figure 4. The viewing behaviour data is the same as in Figure 1. The difference between the figures is that in Figure 4 the vertical dashed lines mark the change points determined by `ruptures` instead of the actual change points checked by hand. From the figure it can be seen, that the start and end times of advertisement breaks are quite accurate, but for the opening and closing credits only one change point was found. The change point for the closing credits seems to align with the beginning of the credits and the change point for the opening credits is in the middle of the credits.



(a) `Pelt` c_{L2} output for Figure 1a data



(b) `Pelt` c_{L2} output for Figure 1b data

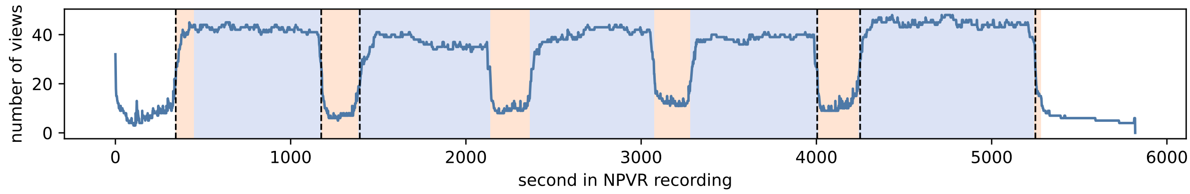
Figure 4: `Pelt` c_{L2} output for Figure 1 data

Trying `Pelt` c_{L2} for Table 1 and 2 recordings, it turns out that the output similar to the above description is the most common result. Typical number of change points in the output can be defined as:

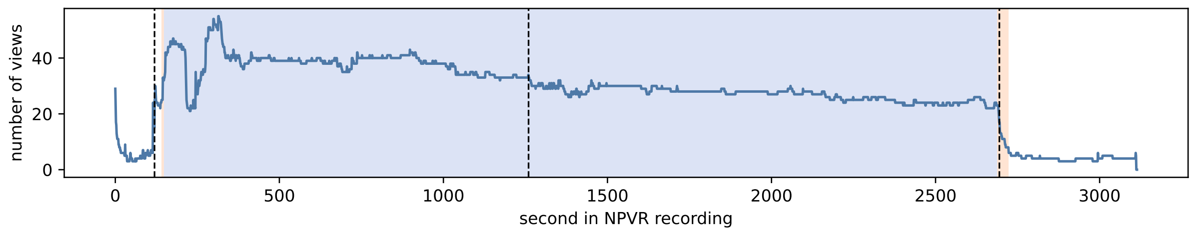
$$k_{typical} = 2 + 2n_{ads} \tag{2}$$

where n_{ads} is the number of advertisement breaks in the recording.

Occasionally the output differs from the default, by the number of change points k being other than $k_{typical}$. An example of this kind of result is visualised in Figure 5. In Figure 5a the change points for the two centremost advertisement breaks were not found. In Figure 5b there is an extra change point in the middle of the core program content. Both example results are worse than the expected result with $k_{typical}$ change points. Nevertheless, it is possible for a result with $k > k_{typical}$ to be better than the defalt result, by having separate change points for the start and end of credits. Unfortunately, there is no trivial method to determine whether an extra change point in `Pelt` output with $k > k_{typical}$ is incorrect, or if it corresponds to an actual change point. In order to make evaluation more straigthforward, `Pelt` output with atypical number of change points will be discarded from further analysis.



(a) Too few change points



(b) Too many change points

Figure 5: `Pelt` c_{L2} output with incorrect number of change points

Change points were calculated with `Pelt` as the search method and c_{L_2} as the cost function for all of the 206 recordings listed in Table 1 and 2. For each recording the change point detection was done with nine different penalty values, ranging from 10000 to 90000, with an increase of 10000 between each value. Listed in Table 4 is the number of recordings $n_{k_{typical}}$ where $k = k_{typical}$ for each penalty. From the table it can be seen that penalty of 70000 produced most conistently the expected number of change points $nk_{typical}$, by having $k = nk_{typical}$ for 201 recordings out of 206 recordings. The results with penalty of 70000 will be discussed in more detail below. Statistics of the result with `Pelt` c_{L_2}, penalty=70000 are listed in Table 5.

penalty	10000	20000	30000	40000	50000	60000	70000	80000	90000
$n_{k_{typical}}$	87	176	193	197	197	200	201	200	195

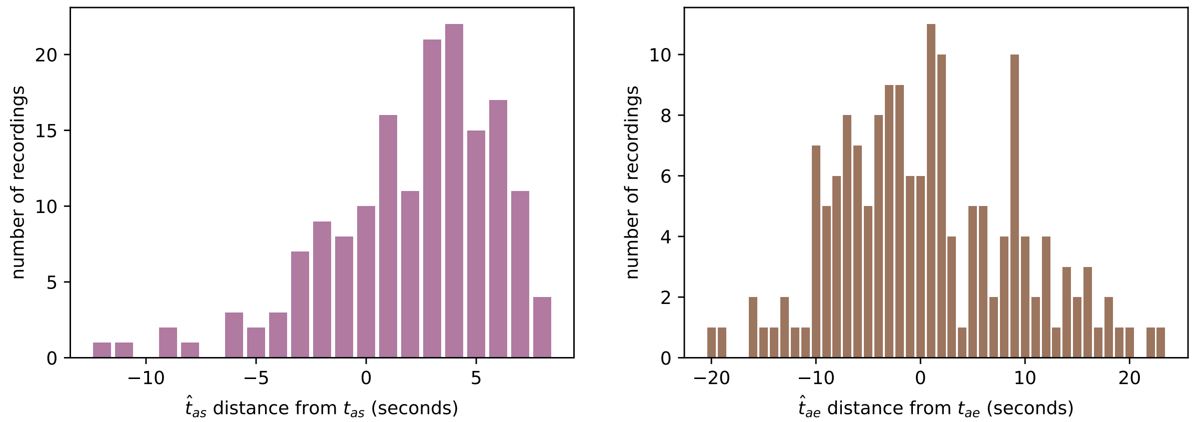
Table 4: Penalty value and the number of recordings with $k_{typical}$ change points

statistic	$\hat{t}_o - t_{os}$	$\hat{t}_o - t_{oe}$	$\hat{t}_c - t_{cs}$	$\hat{t}_o - t_{os}$	$\hat{t}_{as} - t_{as}$	$\hat{t}_{ae} - t_{ae}$
sample size	206	206	206	206	328	328
minimum	-18	-166	2	-107	-12	-20
1st quartile	9	-20	5	-46	0	-6
median	20	-9	7	-30	3	0
3rd quartile	38	-5,25	9	-23	5	7
maximum	157	152	136	14	8	23
variance	497	726	227	896	15,3	75,0
standard deviation	22,3	27,0	15,1	29,9	3,91	8,66

Table 5: Five-number summary and other statistics for `Pelt` c_{L2}, penalty=70000

The five-number summary of the quartiles in Table 5 gives some insight about the output, but the results can be understood more intuitively by plotting the data. Plotted in Figure 6 is the difference between the actual and predicted change points for all of the advertisement breaks in Table 2 recordings. Figure 6a shows the differences in the start time $\hat{t}_{as} - t_{as}$ and Figure 6b shows the differences in the end time $\hat{t}_{ae} - t_{ae}$.

Start of advertisement break $\hat{t}_{as}$ tends to be predicted more accurately than the end $\hat{t}_{ae}$, as the range for $\hat{t}_{as} - t_{as}$ is $\sim \pm 10s$ and it also has lower variance than $\hat{t}_{ae} - t_{ae}$. The range for $\hat{t}_{ae} - t_{ae}$ is $\sim \pm 20s$. It is also worth noting that the distribution for $\hat{t}_{as}$ is a few seconds off from the midpoint $\hat{t}_{as} - t_{as} = 0$, whereas for $\hat{t}_{ac}$ the values are distributed more evenly around $\hat{t}_{ac} - t_{ac} = 0$.



(a) Start of advertisement break

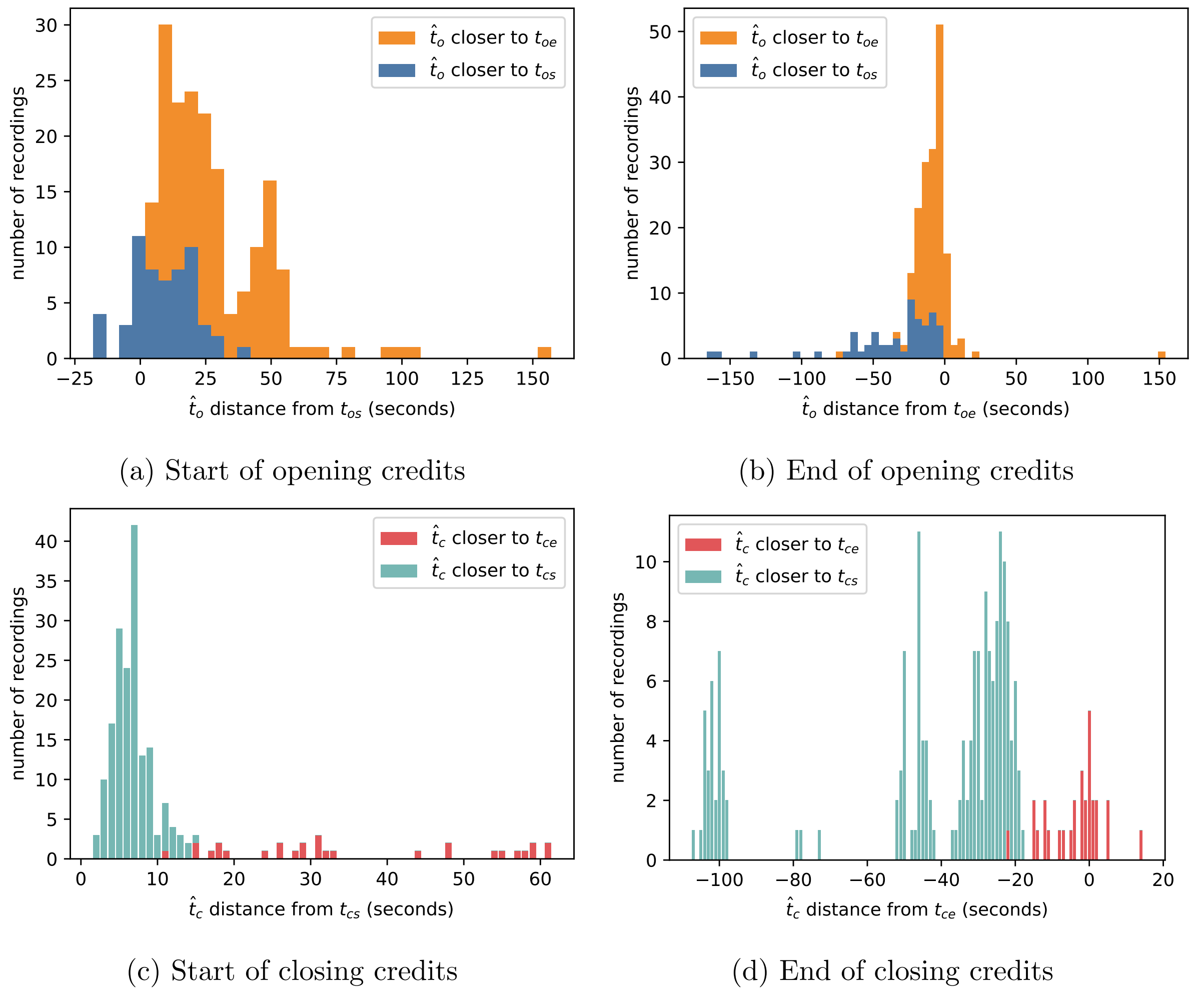
(b) End of advertisement break

Figure 6: `Pelt` c_{L2} output compared to actual change points of advertisement breaks

Plotted in Figure 7 are the predicted locations for opening and closing credits, $\hat{t}_o$ and $\hat{t}_c$, compared to the actual start and end times of the credits. Figure 7a has $\hat{t}_o - t_{os}$ values plotted, Figure 7b has $\hat{t}_o - t_{oc}$ values plotted, et cetera. The results are not as close to the normal distribution as the results for the advertisement breaks in Figure 6.

In order to gain more insight about the nature of the results, the output change points were divided into two populations, depending on whether the predicted change point $\hat{t}$ is closer to the start or the end of the change segment it represents. For the opening credits it seems that $\hat{t}_o$ typically falls in the middle of the credits and aligns closely to neither t_{os} or t_{oc}.

For the closing credits, it is visible in Figure 7c and 7d that most often $\hat{t}_c$ is very close to t_{cs}, although with a delay of few seconds. When $\hat{t}_c$ is more than 10 seconds away from t_{cs} it is typical that it actually aligns to t_{ce} with an accuracy of $\sim \pm 20s$. Another detail worth noting is that although three fourths of the sample $\hat{t}_c$ results are within $9s$ from t_{cs}, for all samples it holds that $\hat{t}_c > t_{cs}$, meaning that $\hat{t}_c$ is never predicted to be before the closing creidts.



(a) Start of opening credits

(b) End of opening credits

(c) Start of closing credits

(d) End of closing credits

Figure 7: `Pelt` c_{L2} output compared to actual change points of credits

Out of the 206 recordings, $\hat{t}_o$ is somewhere on the opening credits in 180 recordings and

$\hat{t}_e$ is somewhere on the closing credits in 199 recordings. That means the approximate location of the opening credits was predicted correctly in 87% of the sample recordings and the approximate location of the closing credits was predicted correctly in 97% of the sample recordings.

6 Discussion

It is non-trivial to find a universal solution that works regardless of the number of advertisemet breaks and the length of the recording. When linear constraint is used, like in `Pelt`, most likely a different penalty value would be optimal for non-commercial and commercial channels. Also better result could maybe be achieved if the penalty value was chosen as a function of the recording length. Another approach to solve the number of change points issue could be combining multiple models, for example Bayesian methods or analog to binary signal processing could be used along `Pelt`.

There seems to be some differences in the results between different kinds of programs. One of the recordings with $k = k_{typical}$ had $\hat{t}_o$ very much off compared to t_o because the recording in question happened to have a slightly unconventional program structure of having opening credits in the middle of the core program content a few minutes in the program. The approach of using user behaviour for change segment detection most likely works best for recordings with a conventional structure with clearly separated interesting and uninteresting content. Structural things that are likely to cause problems for this method are for example opening credits that are not located at the very begining, credits embedded to core program content, recapitulations of previous episodes, sneak peeks of future episodes, especially at the end of closing credits.

Another thing that would benefit from more consideration, is choosing how to clean the data and picking the sample size.

7 Conclusions

The goal of this thesis was to study whether the opening and closing credits and advertisement breaks can be identified from an NVPR recording based solely on user viewing behaviour data. The verdict is that in most cases the general location of an change segment can be found with a reasonable accuracy. However, identifying more exact locations of change points is more difficult. It seems that locating the start of closing credits and the change points for advertisement breaks might be viable, but the end of closing credits and exact start and end for opening credits cannot likely be detected solely from user viewing behaviour.

The accuracy of the results might be improved by considering the differences between different programs and channels. User viewing behaviour is more useful for programs with a typical and predictable structure. If some extra assumptions about the recording structure can be made, it might make it possible to identify non-program content based on user viewing behaviour, despite some change points missing. For example, closing credits length is typically a few minutes at maximum.

For better results it might also be worth considering, if user viewing behaviour based change point detection could be combined with other methods for change point detection.

References

[1] Bellman, R. *On a routing problem.* Quarterly of Applied Mathematics. 1958, vol. 16 :1. P. 87–90. Available at: doi:10.1090/qam/102435.

[2] Berrani, S.-A. & Manson, G. & Lechat, P. *A non-supervised approach for repeated sequence detection in TV broadcast streams.* Signal Processing: Image Communication. 2008, vol. 23:7. P. 525–537. Available at: doi:10.1016/j.image.2008.04.018.

[3] Ibrahim, Z. A. A. & Gros, P. *TV Stream Structuring.* ISRN Signal Processing. 2011, vol. 2011. P. 1–17. Available at: doi:10.5402/2011/975145.

[4] Killick, R. & Fearnhead, P. & Eckley, I. A. *Optimal Detection of Changepoints With a Linear Computational Cost.* Journal of the American Statistical Association. 2012, vol. 107:500. P. 1590–1598. Available at: doi:10.1080/01621459.2012.737745.

[5] Kompatsiaris, Y. & Merialdo, B. & Lian, S. *TV Content Analysis: Techniques and Applications.* CRC Press, 2012. ISBN 978-1-4398-5562-1.

[6] Lévy-Leduc, C. & Roueff, F. *Detection and localization of change-points in high-dimensional network traffic data.* The Annals of Applied Statistics. 2009, vol. 3:2. P. 637–662. Available at: doi:10.1214/08-AOAS232.

[7] Liu, S. & Wright, A. & Hauskrecht, M. *Change-point detection method for clinical decision support system rule monitoring.* Artificial Intelligence in Medicine. 2018, vol. 91. P. 49–56. Available at: doi:10.1016/j.artmed.2018.06.003.

[8] Lung-Yut-Fong, A. & Lévy-Leduc, C. & Cappé, O. *Distributed detection/localization of change-points in high-dimensional network traffic data.* Statistics and Computing. 2012, vol. 22:2. P. 485–496. Available at: doi:10.1007/s11222-011-9240-5.

[9] Manson, G. & Berrani, S.-A. *Automatic TV Broadcast Structuring.* International Journal of Digital Multimedia Broadcasting. 2010, vol. 2010. P. e153160. Available at: doi:10.1155/2010/153160.

[10] Reeves, J. & Chen, J. & Wang, X. L. & Lund, R. & Lu, Q. *A Review and Comparison of Changepoint Detection Techniques for Climate Data.* Journal of Applied Meteorology and Climatology. 2007, vol. 46:6. P. 900–915.

[11] Truong, C. & Oudre, L. & Vayatis, N. *Selective review of offline change point detection methods.* Signal Processing. 2020, vol. 167. P. 107299. Available at: doi:10.1016/j.sigpro.2019.107299.

[12] Tukey, J. W. *Exploratory Data Analysis*, volume 2. 1977. ISBN 0-201-07616-0.

[13] Verbesselt, J. & Hyndman, R. & Newnham, G. & Culvenor, D. *Detecting trend and seasonal changes in satellite image time series*. Remote Sensing of Environment. 2010, vol. 114:1. P. 106–115. Available at: doi:10.1016/j.rse.2009.08.014.

[14] Vert, J.-p. & Bleakley, K. Fast detection of multiple change-points shared by many signals using group LARS. *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.