Indirect Error Difference Approximation for Selecting the Best Anytime Algorithm for Tasks with Unrecognizeable Quality

Konstantin Bulatov
IIAP NAS RA, Yerevan, Armenia
e-mail: hpbuko@gmail.com

https://doi.org/10.62343/csit.2024.7

***Abstract* — In this paper a preliminary study is described of the method of selecting the best anytime algorithm out of two. Given two algorithms which solve the same problem generally improve their results over time, but for which the current quality of the result is now known, the method tries to estimate the difference between their error levels. The method performs its estimation indirectly – first, it constructs a method to approximate the absolute value of the difference between distances from two potential answers to an unknown correct answer, then it uses this approximation to estimate the error value of an anytime algorithm down to an additive constant, and then it uses the latter to compare two algorithms which solve the same problem but observes different intermediate results. The method is tested for the problem of selecting the best strategy for combining per-frame recognition results in the task of text recognition in video stream.**

***Keywords* — Anytime algorithms, combination problems, text recognition, video stream, sequential decision making.**

I. INTRODUCTION

Anytime algorithms are algorithms which could be interrupted at any time and which yield better results between observations [1]. They serve an important role when building intelligent systems, since they allow to use the tradeoff between solution quality and spending a resource (such as computation time) to maximize overall utility of the system.

One of the desired properties of anytime algorithms is *recognizeable quality* [2] – the requiement that the system can observe the current value of the quality of the result. The knowledge of the quality of the current result, or at least how this quality changed between observations, allows to use dynamic programming to solve the problem of opti-mal stopping, or build reward functions for learning optimal behaviour when constructing higher-level decision making schemes. However, there exist a wide class of problems where the quality of the current result is not directly observable, but where the process could be considered as an anytime algorithm with mean results improving over time. The examples of such problems are text recognition in a video stream [3] or sequential tomographic reconstruction with increasing number of available projections, also called "monitored reconstruction" [4]. For such cases the "correct" solution is not known during operation (as in text recognition in video stream), or, strictly speaking, not known at all (in tomographic reconstruction). However, depending on how the problem is stated, there could still exist methods which could allow meaningful decision making relying on the general structure of the problem. For example, in the mentioned problems, if the quality of the result is determined as a distance (calculated as a value of a metric which satisfies triangle inequality) to an "ideal" result, then, for example, an optimal stopping method could be formulated, which only analyses the distances between observed values, and assumes that the algorithm is generally well-behaved (generally improves over time and yields diminishing returns over time).

In this paper we consider the problem of dynamic selection of the best anytime algorithm, given two algorithms which solve the same problem. We construct a method which would not rely on recognizeable quality, and try to approximate the potential difference between the errors of the two observed algorithms.

II. FRAMEWORK

Let Θ respresent the set of all possible solutions for a task with a given function $d : \Theta^2 \to \mathbb{R}$, which have the property $\forall a, b \in \Theta : a = b \Rightarrow d(a, b) = 0$. Let $\theta \in \Theta$ denote an ideal solution of the task. Let us define the quality of the result $r \in \Theta$ as a value inversely related to the value of the function $d(r, \theta)$ (i.e. the lower the value of $d(r, \theta)$ the "closer" the result r is to an ideal result θ). The value $d(r, \theta)$ will be referred to as an "error" of the result r.

Let us consider a family of anytime algorithms which imply the sequence of process stages with numbers $n = 1, 2, \ldots$. Given a certain fixed but unknown θ we observe a sequence of states $S_1, S_2, \ldots, S_n, \ldots$ where $\forall i : S_i = S_i(\theta) \in \mathbb{S}$. Anytime algorithm R on each stage n takes as an input a state S_n, and on each stage n the algorithm could be stopped with a result $R_n(S_1, \ldots, S_n) \in \Theta$. The quality of the result R_n, i.e. $d(R_n(S_1, \ldots, S_n), \theta)$ is unknown during the process, however we will assume that the algorithm R solves a problem of minimizing the total error on some set of input data $D \subset \mathbb{S}^n \times \Theta$:

$$\sum_{(S_1,\ldots,S_n,\theta)\in D} d(R_n(S_1, \ldots S_n), \theta) \to \min_R . \quad (1)$$

Let us now consider two separate anytime algorithms R and Q, both solving the same problem (1). Let us assume that both algorithms during their operation receive the same sequence of input states $S_1, S_2, \dots, S_n$, however their results R_n and Q_n on n-stage could differ. We can state a problem of selecting a result $M_n \in \{R_n, Q_n\}$, $M_n = M_n(R_1, \dots, R_n, Q_1, \dots, Q_n)$ such as to minimize the total error.

Optimal selection of M_n consists of selecting a result with a minimal error, i.e.:

$$M_n = \begin{cases} R_n, & d(R_n, \theta) < d(Q_n, \theta), \\ Q_n, & d(R_n, \theta) \geq d(Q_n, \theta), \end{cases} \tag{2}$$

however in order to solve this problem one needs to estimate the error difference $d(R_n, \theta) - d(Q_n, \theta)$.

The general framework of the approach proposed in this paper consists of three approximations:

1) The approximation of the absolute value of the error difference $|d(a, \theta) - d(b, \theta)|$ for a pair of results $a, b \in \Theta$;
2) Approximation of $d(R_i, \theta)$ except for a additive constant, for a given sequence of anytime algorithm results $R_1, R_2, \dots, R_n$;
3) Approximation of the difference between the additive constants for two sequences of anytime algorithm results, resulting in a hypothesis of selecting the best result out of two.

A. Approximation of the absolute value of the error difference

For two elements $a, b \in \Theta$ instead of trying to estimate $d(a, \theta) - d(b, \theta)$ we will try to approximate the absolute value of this difference $|d(a, \theta) - d(b, \theta)|$. Let us consider a non-negative function $g : \Theta^2 \to \mathbb{R}$ complying to the property $\forall a, b \in \Theta : a = b \Rightarrow g(a, b) = 0$. Let us assume that the value of $g(a, b)$ approximates the absolute value of the difference between error levels of a and b:

$$g(a, b) \approx |d(a, \theta) - d(b, \theta)|. \tag{3}$$

As an example of such approximation, for a certain metric function d we could construct the function $g(a, b)$ as a function proportional to $d(a, b)$ with minimizing a squared approximation error for a given training dataset $D_g \subset \Theta^3$:

$$\begin{gathered} g(a, b) = \alpha \cdot d(a, b) : \\ \sum_{(a,b,\theta) \in D_g} \left(|d(a, \theta) - d(b, \theta)| - \alpha \cdot d(a, b) \right)^2 \to \min_{\alpha \in [0,1]}. \end{gathered} \tag{4}$$

B. Approximation of error except for a constant

Given a sequence of anytime algorithm results $R_1, R_2, \dots, R_n$ for a problem with an ideal solution θ let us construct an approximation of the error $d(R_i, \theta)$ down to an additive constant, i.e. construct $r_1, r_2, \dots, r_n$ such that:

$$\forall i \in \{1, \dots, n\} : \quad \begin{aligned} & d(R_i, \theta) \approx r_i + \gamma, \\ & \text{with } r_n = 0. \end{aligned} \tag{5}$$

Given an approximation of the absolute value of the error differences, let us construct $r_1, \dots, r_n$ by solving the following problem:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(|r_i - r_j| - g(R_i, R_j) \right)^2 \to \min_{r_1, \dots, r_n, \text{ with } r_n = 0}. \tag{6}$$

The problem (6) is easily solvable if for each pair (r_i, r_j) the sign of the different $r_i - r_j$ is known or could be assumed. In general, when anytime algorithms are considered typically it is assumed that the results of the algorithm always improve over time (or at least, in a weaker definitions, "are well-behaved over time"). Thus, if we assume that the results of the algorithm improve over time, we can assume that $r_i - r_j > 0$ for any $i < j$, and thus easily solve the problem (6).

It is also worth to note that if for the approximation of the absolute value of error different we use a linear function of the distance (4), then with any $\alpha > 0$ the estimations $r_1, \dots, r_n$ will be the same except for a multiplicative constant, and if the estimations will be constructed in form $r_i = \alpha \cdot \hat{r}_i$, then simply the distances $d(a, b)$ can be used instead of $g(a, b)$.

C. Approximation of the error difference

Let us consider two sequences of solutions for a problem with an ideal result θ: the sequence $R_1, \dots, R_n$ of the results of the algorithm R and the sequence $Q_1, \dots, Q_n$ of the results of the algorithm Q. Let us assume that the results of both algorithms improve over time, thus by solving the problem (6) we obtain $r_1, \dots, r_n$ and $q_1, \dots, q_n$ such that:

$$\forall i \in \{1, \dots, n\} : \quad \begin{aligned} & d(R_i, \theta) \approx r_i + \gamma_R, \\ & d(Q_i, \theta) \approx q_i + \gamma_Q. \end{aligned} \tag{7}$$

Let $\gamma = \gamma_R - \gamma_Q$ and let us find γ by solving the following problem:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(|r_i - q_j + \gamma| - g(R_i, Q_j) \right)^2 \to \min_{\gamma}. \tag{8}$$

By enumerating all possible combinations of signs under the absolute value bars in (8) we can find all solutions for this problem in polynomial time. Given a found γ we can now make an algorithm selection decision, by selecting the result $M_n \in \{R_n, Q_n\}$ on stage n by supplying the obtained approximation to the original definition (2):

$$M_n = \begin{cases} R_n, & r_n - q_n + \gamma < 0, \\ Q_n, & r_n - q_n + \gamma \geq 0, \end{cases} \tag{9}$$

moreover, if the approximation r_i and q_i was constructed given $r_n = q_n = 0$ then it is enough to simply compare γ with zero.

It is important to note that the task (8) could have many solutions. For example, if $\forall i, j \in \{1, \dots, n\} : R_i = R_j \wedge Q_i = Q_j$ and $g(R_n, Q_n) > 0$ then $\forall i \in \{1, \dots, n\} : r_i = q_i = 0$ and the solution for the problem (8) is $\gamma = \pm g(R_n, Q_n)$, which makes the choice (9) impossible. In order to reduce ambiguity, one can set up a fixed policy to find the maximal value of γ which minimizes (8), thus favouring the algorithm Q in ambigious cases.

As in the previous subsection, it is worth to note that if in both approximations a linear function (4) is used as $g(a, b)$ with the same $\alpha > 0$ then by setting $r_i = \alpha \cdot \hat{r}_1$, $q_i = \alpha \cdot \hat{q}_i$ and $\gamma = \alpha \cdot \hat{\gamma}$ we can find $\hat{\gamma}$ with the described algorithm simply using the distances $d(a, b)$ instead of the function $g(a, b)$.

III. EXPERIMENTAL EVALUATION

For experimental evaluation of the proposed approach let us consider the task of the recognition of text string (text field of a document) in a video stream. Let Θ be a set of all possible text strings, θ – a “correct” text field recognition result. Let the algorithm R denote an algorithm of selecting a single per-frame result with maximal focus estimation score [5], and the algorithm Q represent the modification of the algorithm ROVER [6], [7] to combine per-frame recognition results.

Normalized Levenshtein distance [8] will be used as $d(a, b)$, and we will construct an approximation $g(a, b)$ as a linear function of $d(a, b)$ (4). As was mentioned before, the results of the estimations does not actually depend on α in (4), thus we will simply use the approximation $g(a, b) = d(a, b)$.

As the original dataset the documents from MIDV-500 dataset [9] were used. By recognizing the text fields using the system [10], 691 text field recognition result sequences were obtained. On 491 sequences both the results of R and Q were improving over time on all stages of the process, on 116 – only the results of R, on 72 sequences – only the results of Q, and on the remaining 84 sequences neither the results of R nor Q were strictly improving over time throughout the stages of the process.

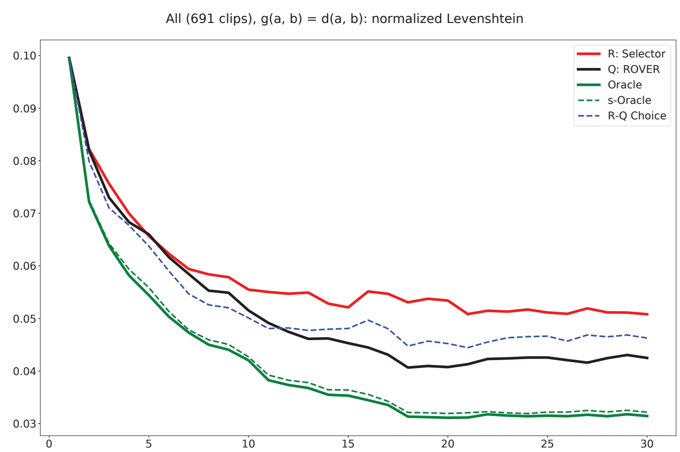


Fig. 1: Error plot for text field recognition in video stream (all sequences). Horizontal axis – frame number, vertical axis – mean error.

Figure 1 shows the dynamics of text field recognition error on all sequences of the dataset. Red and black show the results of the algorithms R and Q respectively, solid green line visualizes an “ideal selection” of the best result (2) (to which a correct result is known). Dashed green line represent the result of the proposed approach with an ideal selection of signs of the differences $r_i - r_j$ in (6). Blue dashed line show the result of the proposed method in a computable scenario (thus, having no access to ideal signs of the differences $r_i - r_j$ and the ideal result), which assumes that R and Q always improve over time.

The selection of the best result using the proposed method allows to reduce mean error (relative to the best of the two algorithms) for sufficiently low amount of frames ($n < 13$). It is worth noting that with ideal selection of signs of the differences $r_i - r_j$ in (6) the selection of the algorithm using the proposed method almost corresponds to an ideal selection.

Figure 2 shows the dynamics of the text field recognition error on the four subsets of the dataset. It can be noted that on the sequences where both R and Q strictly improve over time the proposed method of selecting the best result is almost indistinguishable from ideal choice. The highest error is observed on the subset where only Q (which has the highest overall mean quality) improves over time, as the method R not only does not improve over time on this subset, but also reaches the highest levels of error.

IV. CONCLUSION

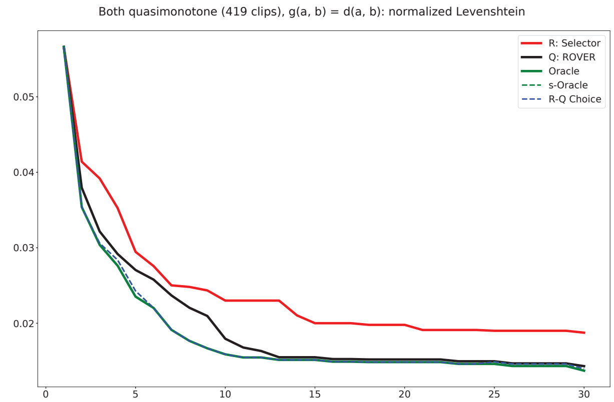
As it can be seen from the experimental evaluation, the proposed method could show promise in some practical applications, however the most crucial problem is deal with the fact that on real data the combination algorithms not always improve over time. If that were always the case, a much simpler method would be to simply select the algorithm which is most different to the known common result (which is the result on the first frame, if the target problem is text field recognition in a video stream).

However, the proposed method achieves almost optimal selection if the sign of the error difference becomes known. Thus, if there exists some other problem-specific method of determining signs of the differences $r_i - r_j$ in (6), then the proposed approach may yield good results.

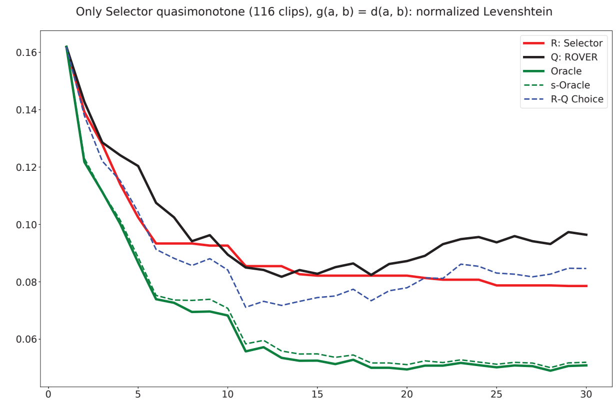
As a function $g(a, b)$ which approximates the absolute value of the error difference another type of function could be used (or trained) which would take not only the two results R_i, R_j or R_i, Q_j, but all the available information $S_1, \ldots, S_n$.

Using the proposed indirect approach it is difficult to solve the problems (6) and (8) simultaneously, since the absolute value bars in (8) has to be eliminated in a different way for different solutions of (6). That is why in the provided general scheme we rely on an assumption of known signs which allow to eliminate absolute value bars in (6) (e.g. by assuming that the algorithms strictly improve over time).

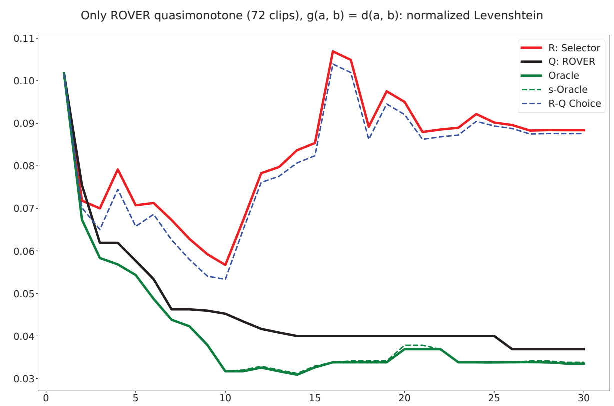
The solution to the problem (6) can also potentially be used to solve the stopping problem for the anytime algorithm R [3], [4]: the optimal stopping implies the estimation of the expected difference $d(R_n, \theta) - d(R_{n+1}, \theta)$, which can try to predict by estimating all $d(R_i, \theta) - d(R_{i+1}, \theta)$ in the form $r_i - r_{i+1}$.



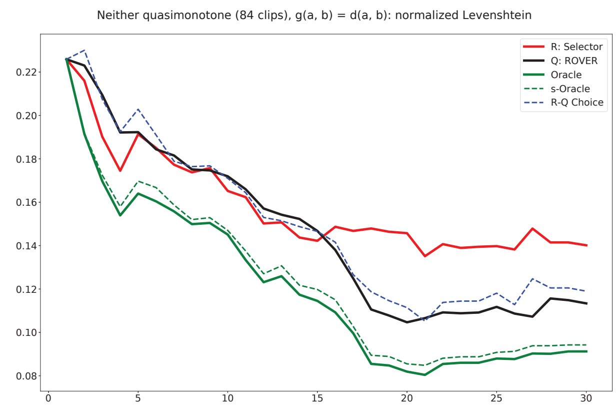
(a) Both R and Q improve over time



(b) Only R improves over time



(c) Only Q improves over time



(d) Neither R nor Q improve over time

Fig. 2: Error plot for text field recognition in video stream (subsets of sequences). Horizontal axis – frame number, vertical axis – mean error.

REFERENCES

1. T. Dean and M. Boddy, "An analysis of time-dependent planning", Proceedings of the Seventh AAAI National Conference on Artificial Intelligence (AAAI'88), pp. 49–54, 1988.
2. S. Zilbershtein, "Using Anytime Algorithms in Intelligent Systems", AI Magazine, vol. 17, no. 3, pp. 73–83, 1996.
3. K. Bulatov, N. Razumnyi and V.V. Arlazarov, "On optimal stopping strategies for text recognition in a video stream as an application of a monotone sequential decision model", IJDAR, vol. 22, no. 3, pp.303–314, 2019.
4. K. Bulatov, M. Chukalina, A. Buzmakov, D. Nikolaev and V.V. Arlazarov, "Monitored Reconstruction: Computed Tomography as an Anytime Algorithm", IEEE Access, vol. 8, pp. 110759–110774, 2020.
5. I. Tolstov, S. Martynov, V. Farsobina and K. Bulatov, "A modification of a stopping method for text recognition in a video stream with best frame selection", Proceedings of ICMV 2020, pp. 116051M1-116051M9, 2021.
6. J.G. Fiscus, "A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER)", Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 347–354, 1997.
7. K. Bulatov, "A Method to Reduce Errors of String Recognition Based on Combination of Several Recognition Results with Per-Character Alternatives", Bulletin of the South Ural State University, Series: Mathematical Modelling, Programming and Computer Software, vol. 12, no. 3, pp. 74–88, 2019.
8. L. Yujian and L. Bo, "A Normalized Levenshtein Distance Metric", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 1091–1095, 2007.
9. V.V. Arlazarov, K. Bulatov, T. Chernov and V.L. Arlazarov, "MIDV-500: A Dataset for Identity Document Analysis and Recognition on Mobile Devices in Video Stream", Computer Optics, vol. 43, no. 5, pp. 818–824, 2019.
10. Y. Chernyshova, A. Sheshkus, V.V. Arlazarov, "Two-step CNN frame-work for text line recognition in camera-captured images", IEEE Access, vol. 8, pp. 32587–32600, 2020.