-SOTA- Document-level Neural Machine Translation

by Lorenzo Lupo

April 2020

Plan

1. Modern Neural Machine Translation
   - Overview
2. Evaluation
   - Automatic metrics
   - Test suites
   - Remarks and conclusions

Plan

1. Modern Neural Machine Translation
   - Overview
2. Evaluation
   - Automatic metrics
   - Test suites
   - Remarks and conclusions

Overview

- MT objective
- from SMT to NMT (attention?)
- sota models
  - transformer
  - transformer variations like Compressive Transformer, Reformer, etc.
- has MT reached human parity? [Lubli et al., 2018]). No, we need DLNMT.
- discourse phenomena, what are they?
- DLNMT objective

Overview

Note: context here is mostly used to indicate the sentences of a document that are not the one currently being translated (both source or target side)

MT output is usually evaluated by **average translation quality** metrics such as BLUE [Papineni et al., 2002] and METEOR [Banerjee and Lavie, 2005]. They are calculate at sentence level by on the base of the number of overlapping n-grams between the translation and the reference. The document-level score is simply an average of the sentence-level scores.

Plan

Evaluation

- Classical metrics such as BLUE and METEOR are inadequate in evaluating document-level MT because they evaluate **average translation quality** at **sentence-level**. Thus:
  - they are unable to capture document-wide phenomena like coherence and cohesion [Wong and Kit, 2012]
  - they are not able to measure improvements over discourse phenomena that affect few words but heavily influence fluency and correctness of the translation [Mller et al., 2018]. E.g. pronomial anaphora.
- Evaluation of **discourse phenomena** can be undertaken with:
  - automatic metrics
  - contrastive test suites

Evaluation

Evaluation Type	Discourse Phenomena	Dependency	Reference
Automatic Metric	Pronouns	Alignments, Pronoun lists	[29]
		Alignments, Pronoun lists	[77]
		English in target (anaphoric)	[43]
	Lexical Cohesion	Lexical cohesion devices	[120]
		Topic model, Lexical chain	[21]
	Discourse Connectives	Alignments, Dictionary	[26]
		Discourse parser	[25, 39]
		Discourse parser	[99]
Test Suites	Pronouns	En→Fr	[23]
		En→Fr (anaphora)	[7]
		En→De (anaphora)	[78]
	Cohesion	En→Fr	[7]
		En→Ru	[115]
	Coherence	En→Fr	[7]
		En↔De, Cs↔De, En→Cs	[117]
		En→Cs	[90]
	Conjunction	En/Fr→De	[85]
	Deixis, Ellipsis	En→Ru	[115]
	Grammatical Phenomena	En→De	[93]
		De→En	[2]
	Word Sense Disambiguation	De→En/Fr	[89, 88]
		En↔De/Fi/Lt/Ru, En→Cs	[86]

Figure: Overview of works on discourse phenomena evaluation in MT [Maruf et al., 2019b].

Evaluation

- The evaluation of discourse phenomena in document-level MT, *desiderata*, and particularly the test suites, should:
  - Provide inter-sentential context[1];
  - Focus on context-dependent cases;
    - E.g., pronominal anaphora cases in which the antecedent is in a previous sentence (context-dependent), instead of being in the same sentence (context-independent).
  - Focus on hard cases.
    - E.g., when translating English to French, **he** is easy whereas **it** is hard to translate because ambiguous.

[1] in the remainder of this presentation, we refer to inter-sentential context simply as context.

Automatic metrics

Accuracy of Pronoun Translation [Miculicich Werlen and Popescu-Belis, 2017]:

- *Compatible languages*: conceived for English to French but it has also been extended to other language pairs.
- *Functioning*:
  - Align source, reference and candidate translation with GIZA++ plus some heuristics;
  - Compare candidate and reference pronouns taking into account **equivalent** pronouns and identical pronouns with **different forms** (target language-specific);
  - E.g. *it is difficult* → *il/ce/c' est difficile*.

Automatic metrics

Pronoun pair-wise ranking [Jwalapuram et al., 2019]

- *Rationale1*: **ranking-based evaluation** measures can achieve higher correlations with human judgments, as rankings are simpler to obtain from humans and to train models on.
- *Compatible languages*: all languages. The metric **only needs target-side inputs** $\implies$ thus it can be trained and evaluated without the need of a parallel corpus for each source-target pair.
- *System input*: a pair $R = (C_r, r)$ and $S = (C_s, s)$ of translations to be compared, where:
  - C_r, C_s are the two translations. Each C can comprise one or multiple sentences (context)
  - r, s are the positions of the pronouns to be compared in the translation R and S, respectively.

Automatic metrics

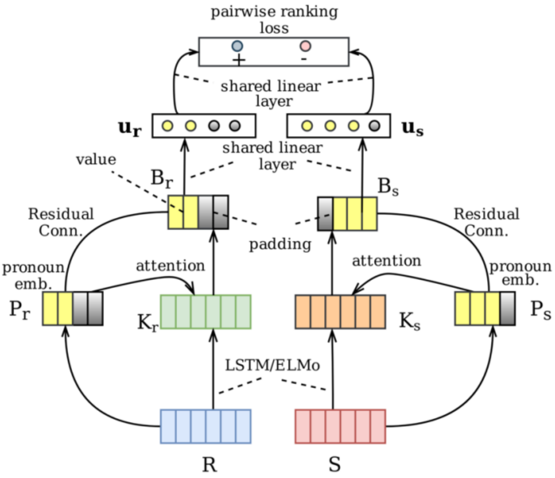

Figure: Pairwise ranking system by [Jwalapuram et al., 2019].

Test suites

- [Bawden et al., 2018]: exemplary contrastive test suite, also good model reaching SOTA. Coherence very bad. Need for good models in coherence?
- [Mller et al., 2018]. Proposal: A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation.
  - Rationale: problems with previous contrastive test suites is that they are either too small to provide stathistical significance [Bawden et al., 2018] or not adapted to properly test DLNMT systems because lemmatized or not always with context.
  - similar method will be adopted by [Jwalapuram et al., 2019]
  - Focus: inter-sentential anaphora, hard case, , i.e., it er, sie, es.

Remarks and conclusions

- automatic metrics
  - are less expensive than human annotation and thus more easily applicable to all languages
  - are noisy because they often rely on other imperfect NLP systems. E.g. alignment and coreference systems.
  - some automatic metrics might not be enough correlated with human judgment and miss the evaluation of some pronominal functions:
    - is the case for APT, for example [Guillou and Hardmeier, 2018]
  - there is nothing on coherence although it's the most relevant for post-editors
- test suites
  - systems trained on in-domain data perform better?
- what could we do?
  - strongly test new automatic metrics against human judgment
  - semi-automatic metrics: use a high precision automatic metric and a human to evaluate negative cases
  - keep designing test suites for very restricted scope

Thank you for your attention!

References I

CiteSeerX A study of translation edit rate with targeted human annotation.
00000.

Agrawal, R. R., Turchi, M., and Negri, M. (2018).
Contextual Handling in Neural Machine Translation: Look Behind, Ahead and on Both Sides.
pages 11–20.
00007 Accepted: 2018-08-08T15:15:28Z.

Banerjee, S. and Lavie, A. (2005).
METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments.
In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

References II

00000.

Barrault, L., Bojar, O., Costa-juss, M. R., Federmann, C., Fishel, M., Graham, Y., Haddow, B., Huck, M., Koehn, P., Malmasi, S., Monz, C., Mller, M., Pal, S., Post, M., and Zampieri, M. (2019).
Findings of the 2019 Conference on Machine Translation (WMT19).
In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 1–61, Florence, Italy. Association for Computational Linguistics.
00000.

References III

Bawden, R., Sennrich, R., Birch, A., and Haddow, B. (2018).
Evaluating Discourse Phenomena in Neural Machine Translation.
In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1304–1313, New Orleans, Louisiana. Association for Computational Linguistics.
00055.

Cao, Q. and Xiong, D. (2018).
Encoding Gated Translation Memory into Neural Machine Translation.
In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3042–3047, Brussels, Belgium. Association for Computational Linguistics.
00005.

References IV

Dobreva, R., Zhou, J., and Bawden, R. (2019).
Document Sub-structure in Neural Machine Translation.
arXiv:1912.06598 [cs].
00000 arXiv: 1912.06598.

Fu, H., Liu, C., and Sun, J. (2019).
Reference Network for Neural Machine Translation.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3002–3012, Florence, Italy. Association for Computational Linguistics.
00000.

References V

Gong, Z., Zhang, M., and Zhou, G. (2015).
Document-Level Machine Translation Evaluation with Gist Consistency and Text Cohesion.
In *Proceedings of the Second Workshop on Discourse in Machine Translation*, pages 33–40, Lisbon, Portugal. Association for Computational Linguistics.
00006.

Guillou, L. and Hardmeier, C. (2018).
Automatic Reference-Based Evaluation of Pronoun Translation Misses the Point.
In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4797–4802, Brussels, Belgium. Association for Computational Linguistics.
00008.

References VI

Jean, S., Bapna, A., and Firat, O. (2019).
Fill in the Blanks: Imputing Missing Sentences for Larger-Context Neural Machine Translation.
arXiv:1910.14075 [cs].
00000 arXiv: 1910.14075.

Jean, S. and Cho, K. (2019).
Context-Aware Learning for Neural Machine Translation.
arXiv:1903.04715 [cs].
00003 arXiv: 1903.04715.

Jean, S., Lauly, S., Firat, O., and Cho, K. (2017).
Does Neural Machine Translation Benefit from Larger Context?
arXiv:1704.05135 [cs, stat].
00038 arXiv: 1704.05135.

References VII

Joty, S., Guzmn, F., Mrquez, L., and Nakov, P. (2014).
DiscoTK: Using Discourse Structure for Machine Translation Evaluation.
In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 402–408, Baltimore, Maryland, USA. Association for Computational Linguistics.
00000.

Junczys-Dowmunt, M. (2019).
Microsoft Translator at WMT 2019: Towards Large-Scale Document-Level Neural Machine Translation.
In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 225–233, Florence, Italy. Association for Computational Linguistics.
00000.

References VIII

Jwalapuram, P., Joty, S., Temnikova, I., and Nakov, P. (2019).
Evaluating Pronominal Anaphora in Machine Translation: An Evaluation Measure and a Test Suite.
arXiv:1909.00131 [cs].
00002 arXiv: 1909.00131.

Kim, Y., Tran, D. T., and Ney, H. (2019).
When and Why is Document-level Context Useful in Neural Machine Translation?
arXiv:1910.00294 [cs].
00001 arXiv: 1910.00294.

References IX

Koehn, P. (2004).
Statistical Significance Tests for Machine Translation Evaluation.
In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395, Barcelona, Spain. Association for Computational Linguistics.
00000.

Kuang, S., Xiong, D., Luo, W., and Zhou, G. (2018).
Modeling Coherence for Neural Machine Translation with Dynamic and Topic Caches.
In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 596–606, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
00012.

References X

- Li, L., Jiang, X., and Liu, Q. (2019).
  Pretrained Language Models for Document-Level Neural Machine Translation.
  arXiv:1911.03110 [cs].
  00001 arXiv: 1911.03110.
- Lubli, S., Sennrich, R., and Volk, M. (2018).
  Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation.
  arXiv:1808.07048 [cs].
  00035 arXiv: 1808.07048.
- Mac, V. and Servan, C. (2019).
  Using Whole Document Context in Neural Machine Translation.
  arXiv:1910.07481 [cs].
  00000 arXiv: 1910.07481.

References XI

Martnez Garcia, E., Creus, C., and Espaa-Bonet, C. (2019).
Context-Aware Neural Machine Translation Decoding.
In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 13–23, Hong Kong, China. Association for Computational Linguistics.
00000.

Maruf, S. and Haffari, G. (2018).
Document Context Neural Machine Translation with Memory Networks.
In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1275–1284, Melbourne, Australia. Association for Computational Linguistics.
00032.

References XII

Maruf, S., Martins, A. F. T., and Haffari, G. (2019a).
Selective Attention for Context-aware Neural Machine Translation.
arXiv:1903.08788 [cs].
00012.

Maruf, S., Saleh, F., and Haffari, G. (2019b).
A Survey on Document-level Machine Translation: Methods and Evaluation.
arXiv:1912.08494 [cs].
00000 arXiv: 1912.08494.

Melamed, I. D., Green, R., and Turian, J. P. (2003).
Precision and Recall of Machine Translation.
In *Companion Volume of the Proceedings of HLT-NAACL 2003 - Short Papers*,
pages 61–63.
00000.

References XIII

Miculicich, L., Ram, D., Pappas, N., and Henderson, J. (2018).
Document-Level Neural Machine Translation with Hierarchical Attention Networks.
arXiv:1809.01576 [cs].
00024 arXiv: 1809.01576.

Miculicich Werlen, L. and Popescu-Belis, A. (2017).
Validation of an Automatic Metric for the Accuracy of Pronoun Translation (APT).
In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 17–25, Copenhagen, Denmark. Association for Computational Linguistics.
00000.

References XIV

Mller, M., Rios, A., Voita, E., and Sennrich, R. (2018).
A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation.
In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 61–72, Brussels, Belgium. Association for Computational Linguistics.
00013.

Ohtani, T., Kamigaito, H., Nagata, M., and Okumura, M. (2019).
Context-aware Neural Machine Translation with Coreference Information.
In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 45–50, Hong Kong, China. Association for Computational Linguistics.
00000.

References XV

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002).
Bleu: a Method for Automatic Evaluation of Machine Translation.
In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
10862.

Popescu-Belis, A. (2019).
Context in Neural Machine Translation: A Review of Models and Evaluations.
arXiv:1901.09115 [cs].
00007 arXiv: 1901.09115.

Post, M. (2018).
A Call for Clarity in Reporting BLEU Scores.
arXiv:1804.08771 [cs].
00100 arXiv: 1804.08771.

References XVI

Scherrer, Y., Tiedemann, J., and Loiciga, S. (2019).
Analysing concatenation approaches to document-level NMT in two different domains.
In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 51–61, Hong Kong, China. Association for Computational Linguistics.
00000.

Stojanovski, D. and Fraser, A. (2018).
Coreference and Coherence in Neural Machine Translation: A Study Using Oracle Experiments.
In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 49–60, Brussels, Belgium. Association for Computational Linguistics.
00003.

References XVII

Sugiyama, A. and Yoshinaga, N. (2019).
Data augmentation using back-translation for context-aware neural machine translation.
In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 35–44, Hong Kong, China. Association for Computational Linguistics.
00000.

Sukhbaatar, S., Grave, E., Bojanowski, P., and Joulin, A. (2019).
Adaptive Attention Span in Transformers.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 331–335, Florence, Italy. Association for Computational Linguistics.
00012.

References XVIII

Tan, X., Zhang, L., Xiong, D., and Zhou, G. (2019).
Hierarchical Modeling of Global Context for Document-Level Neural Machine Translation.
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1576–1585, Hong Kong, China. Association for Computational Linguistics.
00002.

Tiedemann, J. and Scherrer, Y. (2017).
Neural Machine Translation with Extended Context.
In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92, Copenhagen, Denmark. Association for Computational Linguistics.
00038.

References XIX

Tu, Z., Liu, Y., Shi, S., and Zhang, T. (2017).
Learning to Remember Translation History with a Continuous Cache.
arXiv:1711.09367 [cs].
00037 arXiv: 1711.09367.

Voita, E., Sennrich, R., and Titov, I. (2019a).
Context-Aware Monolingual Repair for Neural Machine Translation.
arXiv:1909.01383 [cs].
00003 arXiv: 1909.01383.

Voita, E., Sennrich, R., and Titov, I. (2019b).
When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1198–1212, Florence, Italy. Association for Computational Linguistics.

References XX

00007.

Voita, E., Serdyukov, P., Sennrich, R., and Titov, I. (2018).
Context-Aware Neural Machine Translation Learns Anaphora Resolution.
In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1264–1274, Melbourne, Australia. Association for Computational Linguistics.
00047.

Wang, L., Tu, Z., Way, A., and Liu, Q. (2017).
Exploiting Cross-Sentence Context for Neural Machine Translation.
In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2826–2831, Copenhagen, Denmark. Association for Computational Linguistics.
00045.

Wong, B. T. M. and Kit, C. (2012).
Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level.
In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1060–1068, Jeju Island, Korea. Association for Computational Linguistics.
00044.

Xiong, H., He, Z., Wu, H., and Wang, H. (2018).
Modeling Coherence for Discourse Neural Machine Translation.
arXiv:1811.05683 [cs].
00012.

References XXII

Yang, Z., Zhang, J., Meng, F., Gu, S., Feng, Y., and Zhou, J. (2019).
Enhancing Context Modeling with a Query-Guided Capsule Network for Document-level Translation.
Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 1527–1537.
00001 arXiv: 1909.00564.

Yu, L., Sartran, L., Stokowiec, W., Ling, W., Kong, L., Blunsom, P., and Dyer, C. (2019).
Putting Machine Translation in Context with the Noisy Channel Model.
arXiv:1910.00553 [cs].
00000 arXiv: 1910.00553.

References XXIII

Zhang, J., Luan, H., Sun, M., Zhai, F., Xu, J., Zhang, M., and Liu, Y. (2018).
Improving the Transformer Translation Model with Document-Level Context.
In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 533–542, Brussels, Belgium. Association for Computational Linguistics.
00028.

Zheng, Z., Yue, X., Huang, S., Chen, J., and Birch, A. (2020).
Toward Making the Most of Context in Neural Machine Translation.
arXiv:2002.07982 [cs].
00000 arXiv: 2002.07982.

Markov Decision Processes

Reinforcement Learning

General class of algorithms that allow an agent to learn how to behave in a stochastic and possibly unknown environment by trial-and-error.

Markov Decision Process (MDP)

stochastic dynamical system specified by $< \mathbb{S}, \mathbb{A}, \mathcal{P}, \mathcal{R}, \gamma >$

1. $(\mathbb{S}, \mathcal{S})$ is a measurable state space
2. $(\mathbb{A}, \mathcal{A})$ is a measurable action space
3. $\mathcal{P} : \mathbb{S} \times \mathbb{A} \times \mathcal{S} \to \mathbb{R}$ is a Markov transition kernel
4. $\mathcal{R} : \mathbb{S} \times \mathbb{A} \to \mathbb{R}$ is a reward function
5. $0 < \gamma < 1$ is the discount factor.

Monte-Carlo Policy Gradient: Pseudocode

Input: Stochastic policy π_θ, Initial parameters θ_0, learning rate $\{\alpha_k\}$
Output: Approximation of the optimal policy $\pi_{\theta*} \approx \pi_*$

repeat
  Sample M trajectories $h^{(m)} = \{(s_t^{(m)}, a_t^{(m)}, r_{t+1}^{(m)})\}_{t=0}^{T^{(m)}}$ under policy π_{θ_k}
  Approximate policy gradient

$$\nabla_\theta J(\theta_k) \approx \frac{1}{M} \sum_{m=0}^{M} \sum_{u=0}^{T^{(m)}-1} \nabla_\theta \log \pi_{\theta_k} \left(s_u^{(m)}, a_u^{(m)} \right) \sum_{v \geqslant u}^{T^{(m)}-1} \gamma^{v-u} r_{v+1}^{(m)}$$

  Update parameters using gradient ascent $\theta_{k+1} = \theta_k + \alpha_k \nabla_\theta J(\theta_k)$
  $k \leftarrow k + 1$
until converged

Episodic PGPE Algorithm: Pseudocode

Input: Controller F_θ, hyper-distribution p_ξ, initial guess ξ_0, learning rate $\{\alpha_k\}$
Output: Approximation of the optimal policy $F_{\xi*} \approx \pi_*$

repeat
  for $m = 1, \dots, M$ **do**
    Sample controller parameters $\theta^{(m)} \sim p_{\xi_k}$
    Sample trajectory $h^{(m)} = \{(s_t^{(m)}, a_t^{(m)}, r_{t+1}^{(m)})\}_{t=0}^{T^{(m)}}$ under policy $F_{\theta^{(m)}}$
  end for
  Approximate policy gradient

$$\nabla_\xi J(\xi_k) \approx \frac{1}{M} \sum_{m=1}^{M} \nabla_\xi \log p_\xi \left(\theta^{(m)}\right) \left[G\left(h^{(m)}\right) - b\right]$$

  Update hyperparameters using gradient ascent $\xi_{k+1} = \xi_k + \alpha_k \nabla_\xi J(\xi_k)$
  $k \leftarrow k + 1$
until converged

Truncated Multiple Importance Sampling Estimator

Importance Sampling

Given a bounded function $f : \mathcal{Z} \to \mathbb{R}$, and a set of i.i.d. outcomes $z_1, \ldots, z_N$ sampled from Q, the importance sampling estimator of $\mu := \mathop{\mathbb{E}}_{z \sim P}[f(z)]$ is:

$$\widehat{\mu}_{\mathrm{IS}} = \frac{1}{N} \sum_{i=1}^{N} f(z_i) w_{P/Q}(z_i), \tag{1}$$

which is an unbiased estimator, i.e., $\mathop{\mathbb{E}}_{z_i \overset{\text{iid}}{\sim} Q} [\widehat{\mu}_{IS}] = \mu$.

Truncated Estimator With Balance Heuristic

$$\check{\mu}_{\mathrm{BH}} = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \min \left\{ M, \frac{p(z_{ik})}{\sum_{j=1}^{K} \frac{N_j}{N} q_j(z_{ik})} \right\} f(z_{ik}). \tag{2}$$

OPTIMIST2

Theorem

regretdiscretized Let $\mathcal{X}$ be a d-dimensional compact arm set with $\mathcal{X} \subseteq [-D, D]^d$. For any $\kappa \geqslant 2$, under Assumptions 1 and 2, OPTIMIST2 with confidence schedule $\delta_t = \dfrac{6\delta}{\pi^2 t^2 \left(1 + \lceil t^{1/\kappa} \rceil^d\right)}$ and discretization schedule $\tau_t = \lceil t^{\frac{1}{\kappa}} \rceil$ guarantees, with probability at least $1 - \delta$:

$$\begin{aligned} Regret(T) \leqslant \Delta_0 + C_1 T^{\left(1-\frac{1}{\kappa}\right)} d + C_2 T^{\frac{1}{1+\epsilon}} \\ \cdot \left[v_\epsilon \left((2 + {}^{d}/_{\kappa}) \log T + d \log 2 + \log \frac{\pi^2}{3\delta} \right) \right]^{\frac{\epsilon}{1+\epsilon}}, \end{aligned}$$

where $C_1 = \dfrac{\kappa}{\kappa - 1} LD$, $C_2 = (1+\epsilon)\left(2\sqrt{2} + \dfrac{5}{3}\right) \|f\|_\infty$, and Δ_0 is the instantaneous regret of the initial arm $\boldsymbol{x}_0$.