Incremental Randomized Sketching for Online Kernel Learning

Xiao Zhang Shizhong Liao*

College of Intelligence and Computing, Tianjin University

szliao@tju.edu.cn

June 13, 2019

Outline

Introduction

New Challenges of Online Kernel Learning

(1) High computational complexities

- Per-round time complexity depending on T [Calandriello et al., 2017b]
- Linear space complexity [Calandriello et al., 2017a]

(2) Lack of theoretical guarantees

- Lack of sublinear regrets for randomized sketching [Wang et al., 2016]
- Lack of constant lower bounds on budget/sketch size [Lu et al., 2016]

Main Contribution

Table 1: Comparison with existing online kernel learning approaches (1st order: existing first-order approaches; 2nd order: existing second-order approaches)

	Computational complexities		**Theoretical guarantees**	
	Time (per round)	Space	Budget/Sketch size	Regret
1st order	Constant	Constant	Linear	Sublinear
2nd order	Sublinear	Linear	Logarithmic	Sublinear
Proposed	Constant	Constant	Constant	Sublinear

Main Results

Incremental Randomized Sketching Approach

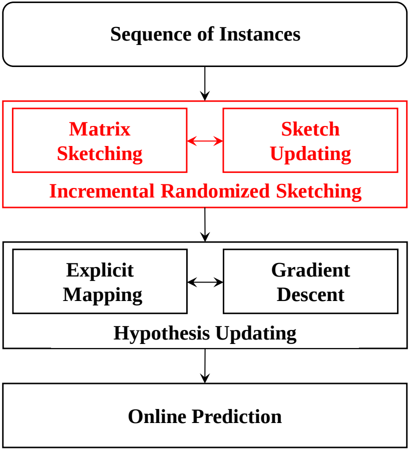

Figure 1: Novel incremental randomized sketching scheme for online kernel learning

Incremental Randomized Sketching Approach

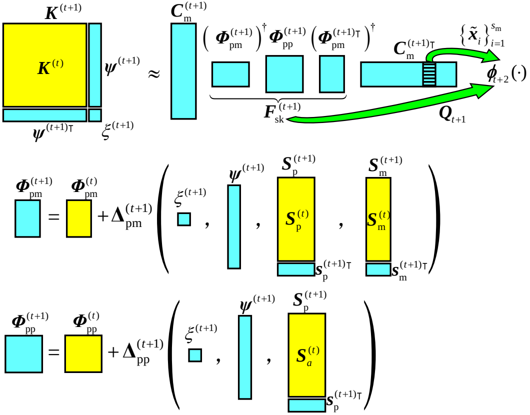

Figure 2: The proposed incremental randomized sketching for kernel matrix approximation at round $t+1$

Main Results

Incremental Randomized Sketching Theory

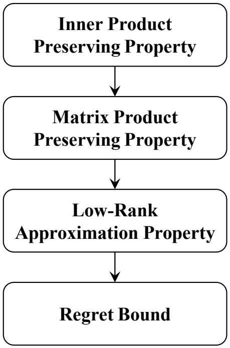

Figure 3: The dependence structure of our theoretical results.

- Product preserving property:
  Statistically unbiased.
- Approximation property:
  $(1+\epsilon)$-relative error bound.
- Regret bound:
  $O(\sqrt{T})$ regret bound,
  constant lower bounds of sketch sizes.

Experimental Results

Table 2: Comparison of online kernel learning algorithms in adversarial environments

Algorithm	german-1		german-2	
	Mistake rate	Time	Mistake rate	Time
FOGD	37.493 ± 0.724	0.140	32.433 ± 0.196	0.265
NOGD	30.918 ± 0.003	0.405	26.737 ± 0.002	0.778
PROS-N-KONS	27.633 ± 0.416	33.984	17.737 ± 0.900	98.873
SkeGD ($\theta = 0.1$)	17.320 ± 0.136	0.329	7.865 ± 0.059	0.597
SkeGD ($\theta = 0.01$)	17.272 ± 0.112	0.402	7.407 ± 0.086	0.633
SkeGD ($\theta = 0.005$)	$\mathbf{16.578 \pm 0.360}$	0.484	7.266 ± 0.065	0.672
SkeGD ($\theta = 0.001$)	16.687 ± 0.155	1.183	$\mathbf{6.835 \pm 0.136}$	1.856

Our incremental randomized sketching achieves a better learning performance in terms of accuracy and efficiency even in adversarial environments.

Conclusion

- Novel incremental randomized sketching for online kernel learning.
- Meet the new challenges of online kernel learning.
  (1) $(1+\epsilon)$-relative error bound.
  (2) Sublinear regret bound under constant lower bounds of the sketch size.
  (3) Constant per-round computational complexities.
- A sketch scheme for both online and offline large-scale kernel learning.

Main References

[Calandriello et al., 2017a] Calandriello, D., Lazaric, A., and Valko, M. (2017a).
Efficient second-order online kernel learning with adaptive embedding.
In *Advances in Neural Information Processing Systems 30*, pages 6140–6150.

[Calandriello et al., 2017b] Calandriello, D., Lazaric, A., and Valko, M. (2017b).
Second-order kernel online convex optimization with adaptive sketching.
In *Proceedings of the 34th International Conference on Machine Learning*, pages 645–653.

[Lu et al., 2016] Lu, J., Hoi, S. C., Wang, J., Zhao, P., and Liu, Z. (2016).
Large scale online kernel learning.
Journal of Machine Learning Research, 17:1613–1655.

[Wang et al., 2016] Wang, S., Zhang, Z., and Zhang, T. (2016).
Towards more efficient SPSD matrix approximation and CUR matrix decomposition.
Journal of Machine Learning Research, 17:1–49.

Thank you!