We study Contextual-Combinatorial Multi-Armed Bandit (CC-MAB) which is tailored for volatile arms and submodular reward functions. CC-MAB inherits properties from both contextual bandit and combinatorial bandit: it aims to select a set of arms in each round based on the side information (context) associated with the arms. By “volatile arms”, we mean that the available arms to select from in each round may change; and by “submodular rewards”, we mean that the total reward achieved by selecting arms is not a simple sum of individual rewards but demonstrates a feature of diminishing returns determined by relations between selected arms (e.g. redundancy). CC-MAB provides an online decision making policy that effectively addresses the issues raised by volatile arms and submodular reward functions. The proposed algorithm is proved to achieve \( O\left(\sqrt{\log(T)}\right) \) regret after a span of \( T \) rounds. The efficacy of CC-MAB is evaluated by experiment conducted on a real-world crowdsourcing dataset, and results show that our algorithm outperforms the prior art.

### Algorithm Design: CC-MAB

- **Context Space Partition**
  - Context space: \( \mathcal{U} = [0,1]^D \), \( D \) is the dimension of context vector.
  - Partition: \( \mathcal{I} = \{ 1,2,\ldots,|\mathcal{I}| \} \) hypercubes of identical size \( \frac{1}{|\mathcal{I}|} \).
  - Each hypercube \( p \in \mathcal{I} \) maintains a counter \( C(p) \) (record the number of times that an arm with \( x \in p \) is chosen), and a quality sample mean \( r(p) \) is calculated.

- **Under-explored arms**
  - \( \mathcal{P} := \{ p \in \mathcal{I} \mid m \in \mathcal{M} \wedge r_m(p) \notin \mathcal{P}\} \)
  - Explore when \( S' = \text{randomly select arms in} \{ m \in \mathcal{M} \mid p_m \in \mathcal{P} \} \)

- **Expand wisely**
  - \( S'' \) = \{ arms \ that \ are \ explored \ in \ the \ previous \ round \}

- **Marginal utility**
  - \( \Delta r(p) = r(p) - \bar{r} \) for arms with context in \( p \).

- **Explore**
  - \( S' \) = \{ arms \ that \ have \ not \ been \ explored \}

- **Exploit**
  - \( S'' \) = { arms \ that \ are \ explored \ in \ the \ previous \ round \}

- **Submodular utility**
  - \( \Delta r(p) = r(p) - \bar{r} \) for arms with context in \( p \).

### Analytical Results

**Hölder Condition**

There exists \( L > 0, a > 0 \) such that for any two contexts \( x, x' \in \mathcal{U} \), it holds that \( |r(x) - r(x')| \leq L|x - x'|^a \).

**Regret Upper Bound**

Let \( K(t) = t^a \log(t) \) and \( R_n = \sqrt{\frac{2}{a} \log n^a} \).

If CC-MAB is run with these parameters and Hölder condition holds true, the regret \( R(T) \) is bounded by

\[
R(T) \leq \sqrt{\frac{2}{a} \log T} \left( 2 \log^2 T + 200 \right) + 160 T^{-a/2} + 32 R_n.
\]

The leading order of the above regret \( R(T) \) is \( \sqrt{T} \). The regret is sublinear.

### Simulation and Discussion

**Setup: Yelp Data**

- **Budget**: \( \mathcal{B} \)
- **Number of arms**: \( |\mathcal{M}| \)
- **Number of context**: \( |\mathcal{I}| \)

**Simulation Results**

**Fig. 1**: Arm quality distribution

**Fig. 2**: Expected quality of hypercubes

**Fig. 3**: Comparison of cumulative rewards

**Fig. 4**: Cumulative rewards over budgets

**Fig. 5**: Regret over budgets

**Fig. 6**: Impact of submodularity