Online Bipartite Matching:
A Survey and A New Problem

Xixuan Feng
xfeng@cs.wisc.edu

Abstract

We study the problem of online bipartite matching, where algorithms have to draw irrevocable matchings based on an incomplete bipartite graph. Specifically, we focus on algorithms that maximize number of matchings (i.e. graphs with weight 0 or 1). First, competitive ratios of a well-studied problem (only one side of the bipartite graph is incomplete) with adversarial, random order and iid models are presented. Second, we discuss how some of these models apply to a similar but rarely-studied problem: online bipartite matching with possibly both sides incomplete.

1 Introduction

Bipartite matching is a fundamental problem in combinatorial algorithm research. In particular, we focus on the problems with online settings, which require algorithms that make optimal decision based on currently available information. Karp and two Vazirani's [6] considered adversarial inputs. They showed straightforward deterministic algorithm GREEDY only has competitive ratio $\frac{1}{2}$, while a simple random algorithm RANKING has an optimal competitive ratio $1 - \frac{1}{e}$. It is still the best result known for adversarial inputs. After first introduced by that, online bipartite matching has drawn much attention because of its many applications. In [5], the competitive ratio is proved to be $1 - \frac{1}{e}$ under random order inputs. Recently, people that don't believe $1 - \frac{1}{e}$ is the bound of the problem have made significant progress based on stochastic (iid model) inputs [4], [1], [7]. Although, most work addresses the problem that one side is fixed and the other side arrives one by one. Full online bipartite matching is the problem that allows both sides arrive in some order. The only result on full online bipartite matching published is in [3], but no algorithm better than competitive ratio $\frac{1}{2}$ is purposed. In this paper, we formulate the problem of full online bipartite matching based on different models and try to build some bases for future breakthrough of the problem.

2 Online Bipartite Matching

In this section, we discuss the problem of online bipartite matching, which has been studied extensively.

2.1 Adversarial Analysis

2.1.1 Problem Statement

In [6], the problem with adversarial inputs are as follows:

Problem 2.1. Given a bipartite graph $G(U, V, E), |U| = |V| = n$ containing a perfect matching, vertices V (the girls) arrive in an order selected by the adversary, and edges incident to a vertex $v \in V$ are unknown by us only until the vertex arrives. As girl v arrives, we may assign a boy $u \in U$ to match her or leave v unmatched forever, and the match is irrevocable. The task is to give a decision sequence that maximize the size of resulting matching.

2.1.2 GREEDY

The most straightforward algorithm is a greedy algorithm that match the first valid boy.

Online Matching
Input v: the new arrival girl; U_{valid}: boys that has edges to v and not picked before
Output u: the boy picked, possible NULL

```
if U_valid is not empty then
    return first boy in U_valid;
else
    return NULL;
```

Theorem 2.1. *If an online algorithm produces a maximal matching upon G, the competitive ratio is at least $\frac{1}{2}$.*

Proof. For every edge (u, v) in the perfect matching of B, either u or v is present in the matching generated by the algorithm. Otherwise, the matching can be augmented by adding (u, v). So, at least $\frac{n}{2}$ vertices in the matching, i.e. it has a competitive ratio at least $\frac{1}{2}$.

GREEDY always produces a maximal matching upon G, so it has a competitive ratio at least $\frac{1}{2}$. But the competitive ratio is actually no better than $\frac{1}{2}$ because the following input forces the algorithm has a matching of size $\frac{n}{2}$. Every vertex in first half (based on arrival order) of V connect to all vertices in U, while the second half only connect to those vertices in U selected by GREEDY, which are known by the adversary. The paper also discussed if the we pick a random valid boy instead of the first one, the competitive ratio become $\frac{1}{2} + O(\frac{log(n)}{n})$, which is no better than GREEDY asymptotically.

2.1.3 RANKING

To achieve a better competitive ratio, a randomized algorithm RANKING is purposed in [6]. The algorithm first assigns a ranking to U before first arrival of girls. Upon each arrival, it matches the girl to the boy with the highest rank.

Ranking

  Randomly permute U and assign the ranking;

Online Matching

Input v: the new arrival girl; U_{valid}: boys that has edges to v and not picked before

Output u: the boy picked, possible NULL

  if U_{valid} is not empty **then**

    return boy in U_{valid} with the highest rank;

  else

    return NULL;

A mistake was first found by Erik Krohn and Kasturi Varadarajan in the proof of Theorem 2.2 presented in [6] 17 years after it published. A correct proof was then given in [5] and [2].

Theorem 2.2. *RANKING with adversarial inputs has the competitive ratio $1 - \frac{1}{e}$, asymptotically.*

2.2 Random Order Analysis

2.2.1 Problem Statement

Problem 2.2. Given a bipartite graph $G(U, V, E), |U| = |V| = n$ containing a perfect matching, vertices V (the girls) arrive in a random order, and edges incident to a vertex $v \in V$ are unknown by us only until the vertex arrives. As girl v arrives, we may assign a boy $u \in U$ to match her or leave v unmatched forever, and the match is irrevocable. The task is to give a decision sequence that maximize the size of resulting matching.

2.2.2 Duality Principle

In [6], a theorem on duality is proved.

Theorem 2.3. *GREEDY with random order inputs and RANKING with adversarial inputs are duals of each other.*

Proof. Base case:
$n = 1$, the matching produced by both algorithms are the same.
Inductive Step:
First we consider the matching produced by RANKING with adversarial inputs, let v denote the highest-ranked boy, it matches to the earliest-arrived girl who has an edge to v, say u. As to random order inputs that boys arrive in the random order as they are ranked in the adversarial case, v arrives first and is matched to u. And the same matching produced in a graph removing v and u, hence, both algorithms produce the same matching.

Therefore, the following theorem is also proved.

Theorem 2.4. *GREEDY with random order inputs has the competitive ratio* $1 - \frac{1}{e}$*, asymptotically.*

2.3 IID Analysis

Many significant results have been discovered since J. Feldman, A. Mehta, V. Mirrokni and S. Muthukrishnan [4]. They are motivated by the application of advertisement display problem.

2.3.1 Problem Statement

Problem 2.3. Given a bipartite graph $G(A, I, E)$ over advertisers A and impression types I, with $|A| = k$ and $|I| = m$. $\forall i \in I$, e_i is the expected number of appearances of i, and $\sum_{i \in I} e_i = n$. Let $\mathcal{D}$ denote the distribution over I defined by $Pr(i) = \frac{e_i}{n}$.
The task is, upon each i.i.d. draws of $i \sim \mathcal{D}$, to decide some advertiser $a \in A$ and $(a, i) \in E$ to match i, or leave i unmatched, in order to maximize the size of resulting matching.

The main different of this problem is that it knows about the edge information of the expected bipartite graph, although the resulting graph is very likely (with probability 1) to be different to the expected bipartite graph. Still, it is natural to get hints from the expected graph.

2.3.2 Suggested Matching

Offline Suggests
Input $G(A, I, E)$: the bipartite graph; $\mathcal{D}$: the distribution to draw i from
Output f_{ai}: the maximal flow, $\forall a \in A, \forall i \in I$, 0 if $(a, i) \notin E$
Construct graph $G^+(\{s\} \bigcup A \bigcup I \bigcup \{t\}, E^+)$ by adding a source s and sink t. $\forall a \in A$, add an edge (s, a) and define the capacity of (s, a) to be 1. $\forall i \in I$, add an edge (i, t) and define the capacity of (i, t) to be e_i. Each edge $(a, i) \in E$ has the capacity 1.
Perform max-flow algorithm and generate f_{ai}, for edge (a, i);
Online Matching
Input i: the new arrival impression; A: the set of advertisers
Output a: the advertiser picked, possible NULL
Draw an element from A, with $Pr(a) = \frac{f_{ai}}{e_i}$.
if a is not picked before **then**
  return a;
else
  return NULL;

The capacity of (s, a) is 1, so $F_a = \sum_i f_{ai} \in \{0, 1\}$. Let $A^* = \{a \in A | F_a = 1\}$. Upon an arrival of i, each $a \in A^*$ has the probability of $\frac{1}{n}$ to be chosen. Therefore, the analysis of this algorithm can be reduced to the classic "balls in bins" problem. And the following Theorem is proved in [4].

Theorem 2.5. *The approximation factor of the suggested matching algorithm is* $1 - \frac{1}{e}$ *with high probability.*

2.3.3 Two Suggested Matching

Offline Suggests

Input $G(A, \hat{I}, \hat{E})$: the expected bipartite graph, where $i \in I$ and edges related are copied e_i times;

Output

Construct graph $G^+(\{s\} \bigcup A \bigcup \hat{I} \bigcup \{t\}, \hat{E}^+)$ by adding a source s and sink t. $\forall a \in A$, add an edge (s, a) and define the capacity of (s, a) to be 2. $\forall i \in \hat{I}$, add an edge (i, t) and define the capacity of (i, t) to be 2. Each edge $(a, i) \in \hat{E}$ has the capacity 1.

Perform max-flow algorithm and generate the set E_f of edges with non-zero flow on them;

Color the cycles and paths in E_f with alternating blue and red. Odd-length paths are colored with more blue than red. Even-length paths started with $a \in A$, alternate blue and red. Even-length paths started with $i \in I$, first two colored blue, and then alternate red and blue.

Online Matching

Input i: the new arrival impression; A: the set of advertisers

Output a: the advertiser picked, possible NULL

  if a is not picked before AND (a, i) is blue **then**

    return a;

  else if a' is not picked before AND (a', i) is red **then**

    return a';

  else

    return NULL;

The first algorithm break through $1 - \frac{1}{e}$ comes from the idea that uses more hints from the expected graph. The intuition is to consider different types of elements in A according to what color edges it is incident to. The possibility are two blue edges, blue and red edges, single blue edge, single red edge or no edge. A competitive ratio $\frac{1-2/e^2}{4/3-2/3e} \approx 0.67$ is achieved.

Theorem 2.6. *The approximation factor of the two suggested matching algorithm is* $0.67 > 1 - \frac{1}{e}$ *with high probability.*

2.3.4 More Improved Results

Although it is a breakthrough result, two suggested matching is for from the end. The idea of using offline hinting is used to achieve higher competitive ratios [1], [7].

3 Full Online Bipartite Matching

In this section, we consider how some of the ideas from online bipartite matching can be borrowed and applied to full online bipartite matching, which has few valuable results known.

3.1 Adversarial Analysis

3.1.1 Problem Statement

Problem 3.1. Given a bipartite graph $G(U, V, E), |U| = |V| = n$ containing a perfect matching, vertices U (the boys) and V (the girls) arrive in an order selected by the adversary, and edges incident to a vertex are unknown by us only until the vertex arrives. As a vertex arrives, we may assign a match of it to an eligible vertex on the other side or leave unmatched (which may be matched later on), and the match, once drawn, is irrevocable. The task is to give a decision sequence that maximize the size of resulting matching.

This problem is harder than online bipartite matching with adversarial inputs, since the latter is former problem with some restrictions on input order. All boys arrive first and girls later.

3.1.2 GREEDY

The GREEDY algorithm still works on this problem. And we introduce the following two theorems.

Theorem 3.1. *GREEDY on the full online bipartite matching with adversarial inputs has the competitive ratio at least* $\frac{1}{2}$.

Proof. Also, it produces a maximal matching, which means it has a competitive ratio at least $\frac{1}{2}$ according to Theorem 2.1.

Theorem 3.2. *GREEDY on the full online bipartite matching with adversarial inputs has the competitive ratio at most* $\frac{1}{2}$.

Proof. Any bipartite graphs with n boys and girls, and a maximal matching of size $\frac{n}{2}$ can be used to construct an adversarial input that forces GREEDY produce that particular maximal matching. The order is, first, to group vertices by pairs, based on the $\frac{n}{2}$-sized matching. Any following arrival of vertices won't produce matching because the existing matching is maximal.

3.1.3 RANKING

Here we discuss why RANKING does not work on the new problem full online bipartite matching.
Similarly, since vertices are known beforehand, a ranking can still be assigned to boys as in RANKING. However, upon an arrival of a girl, not all boys are presented, so we are not sure whether the current highest-ranked eligible boy is the overall highest-ranked eligible one. If we somehow know the answer to this binary question, we could produce the same matching as in one-sided online bipartite matching. An idea that follows this would be making a guess of the answer. If the guess is correct with high probability, an breakthrough algorithm will be invented (better than $\frac{1}{2}$).

3.1.4 Competitive Ratio and Online Restrictions

In full online bipartite matching settings, any matching consists of two decisions, each one associated with one vertex of the matching. Let's denote the first arrival vertex passive and the later one active. An algorithm left the passive vertex unmatched when it arrives, and match the active vertex to the passive one.
Therefore, there is a trade-off between competitive ratio and online restrictions like a vertex may expire if it does not get matched after certain time it arrives.

3.2 Random Order Analysis

3.2.1 Problem Statement

Problem 3.2. Given a bipartite graph $G(U, V, E), |U| = |V| = n$ containing a perfect matching, vertices U (the boys) and V (the girls) arrive in random order, and edges incident to a vertex are unknown by us only until the vertex arrives. As a vertex arrives, we may assign a match of it to an eligible vertex on the other side or leave unmatched (which may be matched later on), and the match, once drawn, is irrevocable. The task is to give a decision sequence that maximize the size of resulting matching.

3.2.2 GREEDY

GREEDY produces the exact same matching in the one-sided online bipartite matching with random order input if all girls arrive in the same order. Because each order (permutation) has the same probability in this model, the competitive ratio remains the same, which is $1 - \frac{1}{e}$.

Theorem 3.3. *GREEDY on the full online bipartite matching with random order inputs has the competitive ratio at least* $1 - \frac{1}{e}$.

Proof. On the problem of one-sided online bipartite matching with random order input, GREEDY has the competitive ratio at least $1 - \frac{1}{e}$ regardless of the permutations of the boys. Therefore, GREEDY on the full online bipartite matching produces matchings of size $n - \frac{n}{e}$, asymptotically, in the worst case. Hence, it has the competitive ratio at least $1 - \frac{1}{e}$.

An idea that is worth noting here is the duality of this result. If we can find a dual with the adversarial inputs, the competitive ratio will be improved to be $1 - \frac{1}{e}$.

3.3 Marketing Clearing Analysis

Motivated by the problem of maximizing number of trades in the applications of marketing clearing, A. Blum, T. Sandholm and M. Zinkevich consider the graph with buyers and sellers, in which edges are defined by valid price and time overlapping. They have no better result than the competitive ratio $\frac{1}{2}$ without subsidize. But subsidize violate the restrictions of irrevocable, which is essential to online algorithms.

4 Conclusion

In this paper, we first focus on a well-studied problem one-sided online bipartite matching. Algorithms based on different input models are discussed separately. A pair of duals represents the state-of-the-art algorithms with adversarial and random order inputs in terms of competitive ratio. A breakthrough result is presented, which makes use of

offline statistics and more following work is continuing.
We then introduced a new problem, full online bipartite matching. Compared to one-sided online bipartite matching, we showed full online bipartite matching is harder to achieve the same competitive ratio. And then we prove that the algorithm GREEDY competitive ratios with adversarial inputs and random order inputs, respectively.

References

[1] Bahman Bahmani and Michael Kapralov. Improved bounds for online stochastic matching. In Mark de Berg and Ulrich Meyer, editors, *ESA (1)*, volume 6346 of *Lecture Notes in Computer Science*, pages 170–181. Springer, 2010.

[2] Benjamin E. Birnbaum and Claire Mathieu. On-line bipartite matching made simple. *SIGACT News*, 39(1):80–87, 2008.

[3] Avrim Blum, Tuomas Sandholm, and Martin Zinkevich. Online algorithms for market clearing. *J. ACM*, 53(5):845–879, 2006.

[4] Jon Feldman, Aranyak Mehta, Vahab S. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *FOCS*, pages 117–126. IEEE Computer Society, 2009.

[5] Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In Shang-Hua Teng, editor, *SODA*, pages 982–991. SIAM, 2008.

[6] Richard M. Karp, Umesh V. Vazirani, and Vijay V. Vazirani. An optimal algorithm for on-line bipartite matching. In *STOC*, pages 352–358. ACM, 1990.

[7] Vahideh H. Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. *CoRR*, abs/1007.1673, 2010.