Secretary Matching

Instructor: Thomas Kesselheim

Today, we will consider the following generalization of the secretary problem. We again have n applicants that arrive online but now we have multiple positions that we hire for. Each applicant i has a qualification score $w(i,j)$ for each job j. We get to know all these scores when the respective candidate arrives. We then have to immediately and irrevocably decide whether we hire her for a position and if so which one or if we reject her. For each position, we may select at most one candidate, and for each candidate we may select at most one position. The objective is to maximize the sum of the scores $w(i,j)$ for which i is assigned to j.

We can formalized this problem as edge-weighted matching (Korula and Pál, [2]).

1. Adversary chooses edge-weighted bipartite graph $G = (L, R, w)$
2. Nature draws permutation of L uniformly
3. Algorithm sees vertices L one after the other with edges and weights; has to select at most one edge; all selected edges must be a matching at all times

Note that here the set of L is the set of online vertices, R is the set of offline vertices.

Like in the first lectures, we will compare $w(\mathrm{ALG}(G))$ and to $w(\mathrm{OPT}(G))$. Let us call an algorithm c-competitive if

$$\mathbf{E}\left[w(\mathrm{ALG}(G))\right] \geq c \cdot w(\mathrm{OPT}(G)) \ ,$$

where the expectation is over the random order of arrival and possibly over internal randomization of the algorithm.

This is a different benchmark than we considered in the last lecture, where we maximized the probability of selecting the best candidate in the sequence. However, also for this benchmark one can show that for the secretary problem we have $c \leq \max_{\tau \in \{0,1,\ldots,n\}} \sum_{t=\tau+1}^{n} \frac{1}{n} \frac{\tau}{t-1}$ for any algorithm. Today, we will get to know an algorithm for which $c \geq \max_{\tau \in \{0,1,\ldots,n\}} \sum_{t=\tau+1}^{n} \frac{1}{n} \frac{\tau}{t-1}$.

1 Algorithm

We will consider the following algorithm (from [1]), which generalizes the threshold algorithm for the secretary problem from last lecture. Again, we do not do anything for the first τ rounds. Afterwards, whenever a new vertex is presented to the algorithm, we compute an optimum solution on the revealed part of the graph. If, in this local solution, the current online vertex is assigned to an unmatched offline vertex, we add this edge to our matching.

2 Analysis

We let $\ell_t \in L$ denote the vertex that arrives in step ℓ and $r_t \in R$ is the associated right-hand side vertex in the tentative edge $e^{(t)} = (\ell_t, r_t)$.

Theorem 7.1. *For Algorithm 1,*

$$\mathbf{E}\left[w(ALG(G))\right] \geq \sum_{t=\tau+1}^{n} \frac{1}{n} \frac{\tau}{t-1} \cdot w(OPT(G)) \ ,$$

Algorithm 1: Bipartite online matching

Input : vertex set R and cardinality $n = |L|$
Output: matching Accept

Let L' be the first τ vertices of L;
for each subsequent vertex $\ell_t \in L - L'$ **do** `// steps` $\tau + 1$ `to` n
  $L' := L' \cup \ell_t$; $t := |L'|$;
  $M^{(t)}$:= optimal matching on $G[L' \cup R]$; `// e.g. by Hungarian method`
  Let $e^{(t)} := (\ell_t, r_t)$ be the edge assigned to ℓ_t in $M^{(t)}$;
  if Accept $\cup\, e^{(t)}$ is a matching **then**
    add $e^{(t)}$ to Accept;

Let OPT be the weight of the optimal matching M^* on the full graph G. We will show that the expected weight of $e^{(t)}$, i.e. of the edge assigned to vertex ℓ_t in the matching $M^{(t)}$, is a significant fraction of OPT. Then, we will analyze the probability of adding this edge to the output matching Accept.

We will think about the random permutation of L being generation by random draws from the set L without replacement. The key idea is to start this process from the end. That is, we first determine ℓ_n, then ℓ_{n-1}, and so on. This already makes clear that, for any fixed choice of $\ell_{t+1}, \ldots, \ell_n$, vertex ℓ_t is drawn uniformly from $L = L' \setminus \{\ell_{t+1}, \ldots, \ell_n\}$. This perspective will help us argue that involved events can be considered as independent.

Lemma 7.2. *The expected weight of the edge $e^{(t)}$ that is tentatively matched in round t is*

$$\mathbf{E}\left[w(e^{(t)})\right] \geq \frac{OPT}{n} \ .$$

Proof. In step t, we have $|L'| = t$ and the algorithm calculates an optimal matching $M^{(t)}$ on $G[L' \cup R]$. The current vertex ℓ_t can be seen as being selected uniformly at random from the set L'. Hence, the expected weight of the edge $e^{(t)}$ is $\mathbf{E}\left[w(e^{(t)}) \mid L'\right] = 1/t \cdot w(M^{(t)})$.

Additionally, $L' \subseteq L$ with size $|L'| = t$ is selected uniformly at random. Therefore the expected value of the matching M^* restricted to $G[L' \cup R]$ is $\mathbf{E}\left[w(M^* \cap (L' \times R))\right] = t/n \cdot \mathrm{OPT}$. Observe that $M^* \cap (L' \times R)$ would be a feasible choice for $M^{(t)}$. So, as $M^{(t)}$ is a maximum-weight matching in $G[L' \cup R]$, its weight is at least the one of $M^* \cap (L' \times R)$. Therefore, its expected value is at least $\mathbf{E}\left[w\left(M^{(t)}\right)\right] \geq t/n \cdot \mathrm{OPT}$. Together this yields the lemma. ☐

Note that the above expectation is only over the random choice of the set L' and the choice of the element to be last in their order. The rest of the proof will exploit the randomness in the order of the remaining $t - 1$ vertices in L'.

Lemma 7.3. *For all $t \geq \tau + 1$ and all choices of $\{u_t, \ldots, u_n\} \subseteq L$, the probability that it is feasible to add edge $e^{(t)}$ to the output matching M is*

$$\mathbf{Pr}\left[e^{(t)} \cup \textsc{Accept} \text{ is a matching} \,\middle|\, \ell_t = u_t, \ldots, \ell_n = u_n\right] \geq \frac{\tau}{t-1} \ .$$

First, we give an intuitive explanation for Lemma 7.3. Observe that, after conditioning on $\ell_t = u_t, \ldots, \ell_n = u_n$, the edge $e^{(t)}$ is not random anymore because $M^{(t)}$ and ℓ_t are determined.

The edge $e^{(t)} = (\ell_t, r_t)$ can only be added to the matching Accept if r_t has not already been matched in an earlier step. Consider the vertex r_t. In any of the preceding steps $k \in$

$\{\tau+1,\ldots,t-1\}$ the vertex r_t was matched only if it was in $e^{(k)}$, i.e. if in $M^{(k)}$ the vertex r_t was assigned to the left-hand side vertex ℓ_k that is $r_k = r_t$.

Again, the last vertex in the order can be seen as being chosen uniformly at random from the k participating vertices on the left-hand side. Hence, the probability of r_t being matched in step k was at most $1/k$. As before, the order of the vertices $1, \ldots, k-1$ is irrelevant for this event. Therefore, also the respective events if some vertex $k' < k$ was matched to r_t can be regarded as independent.

For a formal proof, we show the following proposition inductively. The case $r = r_t$ and $t' = t-1$ then gives the lemma.

Proposition 7.4. *For all $t' \geq \tau+1$, all $r \in R$ and all choices of $\{u_{t'+1},\ldots,u_n\} \subseteq L$, we have*

$$\mathbf{Pr}\left[\bigwedge_{k=\tau+1}^{t'} r \notin e^{(k)} \;\middle|\; \ell_{t'+1} = u_{t'+1},\ldots,\ell_n = u_n\right] \geq \frac{\tau}{t'} \enspace .$$

Proof. We prove this claim by induction on t'. The claim holds for $t' = \tau+1$ because, in this case, we only have to consider the matching $M^{(\tau+1)}$. In this matching, there are at most $\tau+1$ edges and the vertex $v_{\tau+1}$ is chosen uniformly at random from $L^{\tau+1}$. So the probability that r is matched in step $\tau+1$ is $\mathbf{Pr}\left[r \in e^{(\tau+1)}\right] \leq \frac{1}{\tau+1}$.

From now on, we assume that the claim holds for $t'-1$ and we derive that it also holds for t'. Let $\mathcal{E}_{t'}$ denote the event that $(\ell_{t'+1} = u_{t'+1}) \wedge \mathcal{E}_{t'+1}$ for some fixed $u_{t'+1},\ldots,u_n$.

Note that by fixing all arrivals in rounds $t'+1,\ldots,n$, we also fix the matching $M^{(t')}$. Therefore, after conditioning on $\mathcal{E}_{t'}$, $M^{(t')}$ is not random anymore. Let $S \subseteq V \setminus \{u_{t'+1},\ldots,u_n\}$ denote the set of online vertices that are not matched to r in $M^{(t')}$. We now write the probability as the sum of probabilities of disjoint events, depending on which online vertex arrives in round t', as follows

$$\mathbf{Pr}\left[\bigwedge_{k=\tau+1}^{t'} r \notin e^{(k)} \;\middle|\; \mathcal{E}_{t'}\right] = \sum_{u \in S} \mathbf{Pr}\left[\ell_{t'} = u \mid \mathcal{E}_{t'}\right] \mathbf{Pr}\left[\bigwedge_{k=\tau+1}^{t'-1} r \notin e^{(k)} \;\middle|\; \ell_{t'} = u, \mathcal{E}_{t'}\right] \enspace .$$

The size of the set S is at least $|S| \geq t'-1$ because at most one online vertex is matched to r in $M^{(t')}$. Furthermore, for all $u \in V \setminus \{u_{t'+1},\ldots,u_n\}$, we have $\mathbf{Pr}\left[\ell_{t'} = u \mid \mathcal{E}_{t'}\right] = \frac{1}{t'}$ because conditioning on the event $\mathcal{E}$ effectively restricts which set of online vertices arrives in round $1,\ldots,t'$ but not their respective order. Finally, by induction hypothesis, we have

$$\mathbf{Pr}\left[\bigwedge_{k=\tau+1}^{t'-1} r \notin e^{(k)} \;\middle|\; \ell_{t'} = u, \mathcal{E}_{t'}\right] \geq \frac{\tau}{t'-1} \enspace .$$

This way, we get

$$\mathbf{Pr}\left[\bigwedge_{k=\tau+1}^{t'} r \notin e^{(k)} \;\middle|\; \mathcal{E}_{t'}\right] \geq (t'-1)\frac{1}{t'}\frac{\tau}{t'-1} = \frac{\tau}{t'} \enspace . \qquad \square$$

Proof of Theorem 7.1. The weight of the matchting Accept is the sum of the weights of the tentative edges combined with the probability that these edges are feasible. Let A_t be a random variable that is 1 if $\textsc{Accept} \cup e^{(t)}$ is a matching and 0 otherwise. We get

$$\mathbf{E}\left[w(\textsc{Accept})\right] = \mathbf{E}\left[\sum_{t=\tau+1}^{n} w(e^{(t)}) \cdot A_t\right] \enspace .$$

The expected value of the edge $e^{(t)}$ only depends on the set of vertices that have arrived by step t, but not on the order. Whereas A_t depends on the exact ordering of the vertices $v_{\tau+1}, \ldots, v_{t-1}$. We have

$$\mathbf{E}\left[w(\textsc{Accept})\right] = \sum_{t=\tau+1}^{n} \mathbf{E}\left[w(e^{(t)})\right] \cdot \mathbf{Pr}\left[A_t \mid \mathcal{E}_t\right] \quad .$$

Like in the previous proof, the condition on $\mathcal{E}_t = (\ell_{t+1} = u_{t+1}) \wedge \mathcal{E}_{t+1}$ for some fixed $u_{t+1}, \ldots, u_n$ fixes the whole input ordering in the future rounds $t+1$ to n. In addition, $\mathcal{E}_t$ also fixes the set of items that has already arrived up to and including round t.

We evaluate this sum starting with the last index $t = n$. In this way, the randomness used for every index $t \in \{\tau+1, \ldots, n\}$ is exactly the random draw of the item that arrives in round t from the set of all items that previously arrived. This is independent and gives us,

$$\mathbf{E}\left[w(\textsc{Accept})\right] \geq \sum_{t=\tau+1}^{n} \frac{\tau}{t-1} \cdot \frac{\mathrm{OPT}}{n} = \frac{\tau}{n} \cdot \sum_{t=\tau}^{n-1} \frac{1}{t} \cdot \mathrm{OPT} \quad .$$

□

For $\tau = \lfloor n/e \rfloor$, we have $\frac{\tau}{n} \geq \frac{1}{\mathrm{e}} - \frac{1}{n}$ and $\sum_{t=\tau}^{n-1} \frac{1}{t} \geq \ln\left(\frac{n}{\tau}\right) \geq 1$ which gives,

$$\mathbf{E}\left[w(\textsc{Accept})\right] \geq \frac{\tau}{n} \cdot \sum_{t=\tau}^{n-1} \frac{1}{t} \cdot \mathrm{OPT} \geq \left(\frac{1}{\mathrm{e}} - \frac{1}{n}\right) \cdot \mathrm{OPT} \quad .$$

References

[1] Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In Hans L. Bodlaender and Giuseppe F. Italiano, editors, *Algorithms - ESA 2013 - 21st Annual European Symposium, Sophia Antipolis, France, September 2-4, 2013. Proceedings*, volume 8125 of *Lecture Notes in Computer Science*, pages 589–600. Springer, 2013.

[2] Nitish Korula and Martin Pál. Algorithms for secretary problems on graphs and hypergraphs. In Susanne Albers, Alberto Marchetti-Spaccamela, Yossi Matias, Sotiris E. Nikoletseas, and Wolfgang Thomas, editors, *Automata, Languages and Programming, 36th Internatilonal Colloquium, ICALP 2009, Rhodes, Greece, July 5-12, 2009, Proceedings, Part II*, volume 5556 of *Lecture Notes in Computer Science*, pages 508–520. Springer, 2009.