Solving the weighted bipartite matching problem (with contextualizations)

Vidmantas Zemleris

October 14, 2013

Min-cost bipartite matching: given N jobs and M people and a NxM matrix of costs of performing these jobs. Each job has to be assigned to *one* of M people, while one person can perform maximum one job. Assume $M \leqslant N$ (or matrix can be rotated). Another interpretation can be assigning each of N keywords a tag from M tags available (and we have a likelihood matrix). This can be efficiently solved by a well-known Munkres[1] algorithm in $\Theta(N^2M)$.

Contextualization adds additional interdependencies between solution costs, e.g. the person A would agree to perform the job X cheaper, if person be is assigned job Y (e.g. as he wants to be nearby). In the context of keyword search: the tag_j of a keyword kw_i is more likely if it's nearby a related tag_y.

Comparison of known approaches

Notation: N keywords; M is total number of possible tags; $\widetilde{M}$ is the average # of possible tags; k - # of top-k results to return.

Method	**Advantages**	**Disadvantages**
Exhaustive search	▷ easy pruning and contextualization ▷ optimal answers	**slow**, $O(\widetilde{M}^N)$
Munkres[3] gives one best solution in $\Theta(N^2M)$.	▷ quite fast	▷ no contextualization ▷ only one best result
Keymantic[2] - Munkres modified for contextualizations recursively evaluate[2] all mappings pruning on the cost. do contextualization inside Munkres	▷ quite fast ▷ some contextualization ▷ some of top-k answers	▷ **approximate - not global top-k** ▷ **correctness unproven** ▷ no guarantee of all contextualizations
Murty[6] + *Munkres* - top-k matchings to get each additional result, call Munkres to solve $n-1$ smaller assignments of sizes $2..n-1$. Heuristics can greatly improve expected run time[4]	▷ *top-k optimal* solutions ▷ quite fast	▷ no early pruning (partial matching may change a lot) ▷ **no contextualization** [3]
HMM[1] + *List Viterbi*[7] (slightly different model) can start with estimated HMM params: transition probs from contextualizations, output probs from cost matrix	▷ *top-k optimal* solutions ▷ contextualization of *limited length* ▷ quite fast	▷ no pruning ▷ **a tag may get selected many times**
Murty + *Dynamic Munkres* ***(good if few dependencies)*** 1) enumerate over all contextualization possibilities[4] 2) use Murty's to get top-k results over contextualized cost-matrix reusing older sub-solutions[5] costing only $\Theta(NM)$ per modified "line".	▷ **top-k optimal results** ▷ **fast** if # contextualizations is small	▷ exponential for complex contextualizations

Table 1: comparison of different methods

[1] *Munkres* splits the assignment problem into easier ones: 1) maintaining a set of constraints that restrict the currently allowed matches (edges) to be cheap enough, and 2) solving N unweighted bipartite assignments: starting with an empty matching, find an augmenting path to increase the size of matching - new edges are selected or existing deselected; if no augmenting path exist, loosen the constraints on weights.

[2] by solving multiple weighted matchings with Munkres (and modifying the matrices to (not) to choose specific matches)

[3] could do same unproven contextualization within Munkres as in Keymantic; would at least guarantee top-k with limited contextualization

[4] exploring contextualizaions in depth-first order cost-matrix modifications can be reused

References

[1] S. Bergamaschi, F. Guerra, S. Rota, and Y. Velegrakis. A hidden markov model approach to keyword-based search over relational databases. *Conceptual Modeling–ER 2011*, pages 411–420, 2011.

[2] Sonia Bergamaschi, Elton Domnori, Francesco Guerra, Raquel Trillo Lado, and Yannis Velegrakis. Keyword search over relational databases: a metadata approach. In *Proceedings of the 2011 international conference on Management of data*, pages 565–576. ACM, 2011.

[3] Francois Bourgeois and Jean-Claude Lassalle. An extension of the munkres algorithm for the assignment problem to rectangular matrices. *Communications of the ACM*, 14(12):802–804, 1971.

[4] Matt L Miller, Harold S Stone, and Ingemar J Cox. Optimizing murty's ranked assignment method. *Aerospace and Electronic Systems, IEEE Transactions on*, 33(3):851–862, 1997.

[5] G Ayorkor Mills-Tettey, Anthony Stentz, and M Bernardine Dias. The dynamic hungarian algorithm for the assignment problem with changing costs. 2007.

[6] Katta G Murty. Letter to the editor—an algorithm for ranking all the assignments in order of increasing cost. *Operations Research*, 16(3):682–687, 1968.

[7] N. Seshadri and C.E.W. Sundberg. List viterbi decoding algorithms with applications. *Communications, IEEE Transactions on*, 42(234):313–323, 1994.