Solving weighted bipartite assignment problem (with contextualizations)

October 14, 2013

Min-cost bipartite assignment: given N jobs and M people and a NxM matrix of costs of perform these jobs. Each job has to be assigned to *one* of M people, while one person can perform maximum one job. Assume M<=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 well-known problem is efficiently solved by Munkres/Hungarian 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.

1 Comparison of existing approaches

Notation: N keywords; M is total number of possible tags; $\bar{M}$is average # of possible tags

Method	Advantages	Disadvantages
Exhaustive search	easy pruning and contextualization optimal answers	slow, $O(\bar{M}^N)$
Munkres[3] (rectangular version) gets the best solution to assignment problem in $\Theta(N^2M)$.	quite fast	no contextualization only one best result
Keymantic[2] recursively evaluate all mappings[1] pruning on the cost so far. do the contextualization inside Munkres itself	quite fast some contextualization some of top-k answers	only approximate - not all of top-k answers no guarantee all conceptualizations to be explored if not exploring ALL solution space (i.e. not real top-K) seems there's no 100% correctness & termination
Murty[6] + Munkres to get each additional result, call Munkres to solve $n-1$ smaller assignments of sizes $2..n-1$. Can be greatly optimized through heuristics[4]	*top-k optimal* solutions quite fast	no early pruning (an augmenting path may change much...) no contextualization (or at least hard to get)
HMM[1] + List Viterbi[7] or some related conditional models such as CRF	optimal top-k results allows some contextualization of l*imited length* quite fast	no pruning same tag may get selected multiple times
Proposed by us (if few dependencies) 1) enumerate over all contextualization possibilities 2) use Murty to get top-k results over matrix with contextualizations applied 3) reuse older sub-solutions by using the "Dynamic Munkres"[5], so each modified row/column costs $\Theta(NM)$; each modification can be reused if exploring them in depth-first fashion.	**top-k optimal results** **fast** if # contextualizations is small	intractable/exponential for complex contextualizations

Table 1: comparison of different methods

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.