A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance

Memoona Khanum [a]
Tahira Mahboob [b]

[a]Assistant Professor
[a]Department of Computer Science

[b]Assistant Professor
[b]Department of Software Engineering

Warda Imtiaz [c]   Humaraia Abdul Ghafoor [d]   Rabeea Sehar [e]

[b, c, d, e]Department of Software Engineering
[a, b, c, d, e]Fatima Jinnah Women University, Pakistan.

ABSTRACT

The paper is comprehensive survey of methodologies and techniques used for Unsupervised Machine Learning that are used for learn complex, highly non-linear models with millions parameters to used large amount of unlabeled data. Deep belief networks (DBNs) and sparse coding are the two well known techniques of unsupervised learning models. Data clustering distinguishes by the absence of category information. Basically structure in data is finding in clustering and it has long history in scientific field .K-means is the most popular and simple clustering algorithm. This Algorithm was published in 1955. Hierarchical matching pursuit (HMP) for RGB-D data is discussed. Sparse coding learns hierarchical feature representations from raw RGB-D data in an unsupervised way by using hierarchical matching pursuit. The formal study of learning systems is deduced from *Machine learning;* which is a field of research. It has found to be highly interdisciplinary field which acquires and constructs upon ideas from statistics, computer science (engineering), optimization theory, and numerous other disciplines of science and mathematics.

Keywords: Clustering, Feature Selection, Unsupervised Learning, Expectation-Maximization

1. INTRODUCTION

In software engineering automation The research paper provides a tutorial and overview of the field of unsupervised learning in perspective of statistical modeling. Unsupervised learning techniques can be motivated from; information theoretic and Bayesian principles. Basic models in unsupervised learning; including factor analysis, state-space models, some mixtures of Gaussian, hidden Markov models, ICA, PCA, and many variants plus extensions are briefly reviewed here. The aim of this work is the use of unsupervised machine learning to build high-level, class specific feature detectors from unlabeled images and data sets. This approach is influenced by the neuro-scientific conjecture: there exist highly class-specific neurons in the human brain, informally known as "grandmother neurons". The role of labeled data to obtain these class-specific feature detectors is typically emphasized by Contemporary computer vision methodology. Unsupervised algorithms attempt to solve any particular task in harder way than is necessary or, even the

wrong problem altogether is an often made criticism. In the researcher's view, LLE removes much of the force behind this argument because it belongs to a new class of unsupervised learning algorithms. One of the traditional clustering approaches that is K-mean is being offered in the recently developed latent class analysis and related software in which continuous variable are included due to which model base alternative is offered. Different approaches are used in this tool using pre-processing techniques with the integration of evolutionary machine learning algorithms approaches like Pittsburgh and Michigan etc. it allows engineers to perform a very clear and complete analysis of any learning model in comparison to already existing learning model software tools.

2. UNSUPERVISED MACHINE LEARNING TECHNIQUES

2.1) Building Domain Specific Search Engines with Machine Learning Techniques (Andrew McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore)

With the general, Web-wide search engines not possible to increased accuracy and extra functionality but domain specific search engine provides these functionalities and growing in popularity now-a-days. Like www.campsearch.com this search engine provides complex queries, size, location and cost. In this research paper such type of machine learning techniques are used that provide the creation and maintenance of domain specific search engine. We built a demonstration system by using these machine learning techniques. A public hierarchy can be create automatically by unsupervised clustering and can also be search on the base of key words[1]. By using these unsupervised techniques allow creating such type of search engine that search quickly with minimal effort and will be reuse for many domains. A hierarchical organization of materials is provided by search engines like yahoo. In this paper describe new research in information extraction and text classification that enables to provide efficient spidering, identify informative and related text.

2.2) Large-scale Deep Unsupervised Learning using Graphics Processors (Rajat Raina, Anand Madhavan, Andrew Y. Ng)

Unsupervised learning techniques are used for learn complex, highly non-linear models with millions parameters to used large amount of unlabeled data. Deep belief networks (DBNs) and sparse coding are the two well known techniques of unsupervised learning models. For large scale application

these techniques are too slow so these techniques are focusing on smaller scale models. Massively parallel methods are used to resolve these types of problems[2]. Modern graphics processors far surpass the computational capabilities of multi-core CPUs.Modern graphic processors have the potential to modernize the applicability of unsupervised learning methods. Using graphics processors develop general principles for massively parallelizing un-supervised learning tasks .For both DBNs and sparse coding these principles are applied for scaling and learning algorithms. Implementation of Deep belief networks learning is greater than 70 times faster than a dual-core CPU implementation for large models. A simple, inherently parallel algorithm is developed for sparse coding, that have an advantage to a 5 to 15-fold speedup over previous methods.

2.3) Data clustering: 50 years beyond K-means (Anil K. Jain)

One of the most fundamental modes of understanding and learning is organizing data into sensible grouping. The formal study of methods and algorithms for grouping, or clustering, objects according to similarities and same characteristics is Cluster analysis. It does not use category labels that tag objects with prior identifiers, i.e., class labels. Data clustering distinguishes by the absence of category information. Basically structure in data is finding in clustering and it has long history in scientific field[3] .K-means is the most popular and simple clustering algorithm. This Algorithm was published in 1955.In this paper a brief overview of clustering, summarize clustering methods, discuss and elaborate the major challenges and major issues in designing clustering algorithms, and point out these emerging and useful research directions, including semi-supervised clustering, consecutive feature selection during data clustering, and wide-ranging data clustering.

2.4) Unsupervised Feature Learning for RGB-D Based Object Recognition (Liefeng Bo, Xiaofeng Ren , and Dieter Fox)

High quality synchronized videos for both color and depth are provided by RGB-D cameras. With its higher sensing capabilities; it represents a chance to increase the capabilities of object recognition. I the problem of developing features for the color and depth channels of these sensors also increase. In this paper hierarchical matching pursuit (HMP) for RGB-D data are discuss. Sparse coding learns hierarchical feature representations from raw RGB-D data in an unsupervised way by using hierarchical matching pursuit[4]. Wide experiments on various datasets show that using linear support vector machines the features learned with this approach allow superior object detection results. These results are promoting, indicating that current recognition systems may be improved without resorting to careful, manual feature design. The architecture of HMP is manually designed. Automatically learning such structure is very challenging and interesting.

2.5) A Machine Learning Approach to Building Domain-Specific Search Engines(A.Mccallum, K.Nigam, J.Rennie, and K.Seymore)

Domain specific search engines have become popular because of their incredibly increase in extra-features and most importantly accuracy which is not possible with general Web wide Search engine. To some extant there is drawback of domain specific engines that they are time-consuming and difficult in maintenance. Machine learning techniques are proposed for automation of creation and maintenance like Ra Project. It is a technique by which search engine can be created quickly, like with the minimaleffort and in less time[5]. Topics directing spreading is the main focus like substrings extraction which is topics relevant and building the

hierarchy of brows-able topics. Classifier's burden is decreased by using the data which is unlabeled like some class hierarchies and some keywords. In the technique, instead of that builder handle label training data, builder provides the set of few keywords for each category which is useful technique and data can be called labeled data as a rule list classifier.

2.6) Latent class models for clustering: A comparison with K-means(J.Magidson, J.K. Vermunt)

One of the traditional clustering approaches that is K-mean is being offered in the recently developed latent class analysis and related software in which continuous variable are included due to which model base alternative is offered. Two approaches are compared by data simulations in which true members are recognizable. Settings are chosen which are favorable to K-means by simulation and this simulation is done on the basis of assumptions in both techniques K-means and discriminant analysis. In general data which is belongs to true group member is not used in clustering techniques but in discriminant analysis. Discriminant analysis[6] first done on the data set and then it is obtained as a gold standard which shows the upper bound likewise in clustering techniques. By using this technique most surprising results are found that the latent class performance has become so good and is indistinguishable from the actual performance if discriminant analysis.

2.7) KEEL: a software tool to assess evolutionary algorithms for data mining problems(J. Alcalá-Fdez • L. Sánchez • S. García • M. J. del Jesus)

KEEL is software that is used to assess the algorithms which are evolutionary for the field of data mining. It solves problem of data mining like unsupervised learning, classification techniques, and regression of data etc. Different approaches are used in this tool using pre-processing techniques with the integration of evolutionary machine learning algorithms approaches like Pittsburgh and Michigan etc. it allows engineers to perform a very clear and complete analysis of any learning model in comparison to already existing learning model software tools. Basically this software is designed for research and educational purpose. Advance features includes some criteria which are less common in existing software tools like Post-processing, Meta-learning, Statistical tests. KEEL consists of functional blocks like Data management, offline module is design of experiment, and online module is educational experiments.

2.8) Unsupervised Learning by Probabilistic Latent Semantic Analysis (T.HOFMANN)

Novel statistics method which is closely related to latent semantic technique/analysis is described for factor analysis of binary and data count. Contrary to this, there is another method according to which prows from the linear algebras and it performs a Singular Value that has been obtained from data is Decomposed of co-occurrence tables, while the according to proposed technique generative latent class model is used to perform the function of probabilistic mixture decomposition on the data set. Result which is extracted from it is more principled approach according to the solid foundation with the statistical in reference[8]. Proposed technique is to making use of the controlled version regarding temperature. It produces the maximization of algorithms for model fitting, as a result of which it has shown a very good result in practice. There are many applications of probabilistic latent semantic analysis like natural language processing, machine learning from text, information retrieval, and application which uses machine learning techniques.

2.9) Feature Selection for Unsupervised Learning (J. G. Dy and C. E. Brodley)

In this paper, two issues are identified that involves in developing an automated feature subset selection algorithm for unlabeled data; the need [1] for finding the number of clusters in conjunction with feature selection, and [2] normalizing the bias of feature selection criteria in accordance to dimension. Feature Subset Selection using[9] Expectation-Maximization (FSSEM) EM clustering and through two different performance criteria for evaluating candidate feature subsets: scatter separability and maximum likelihood. Proofs on the dimensionality biases of these feature criteria, and a cross-projection normalization scheme that can be applied to any criterion to ameliorate these biases is present. Thus, a normalization scheme is required for the chosen feature selection criterion. The proposed cross-projection criterion normalization scheme was capable of eliminating these biases. Although the research paper under consideration examined: the wrapper framework using FSSEM, feature selection criteria, the search method, and the feature normalization scheme easily applicable to any clustering method. The issues that have been encountered here and solutions presented can be applicable to any feature subset wrapper approach. FSSEM serves as an example; depending on one's application, one may select to apply a more appropriate clustering, search method, and feature selection criteria.

2.10) Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds (L. K. Saul and S. T. Roweis)

Machine learning dilemma commonly begins with the preprocessing of raw multidimensional signals, such as images of faces or spectrograms of speech. To obtain more helpful representations of the information in these signals for subsequent operations such as classification, interpolation, visualization, de-noising, or outlier detection is the aim of preprocessing. If the prior information is missing, above mentioned representations must be discovered automatically. This general framework of unsupervised learning can be used to study automatic methods, which locate unlabelled structure from the statistical regularities of large data sets (Hinton and Sejnowski, 1999)[10].

LLE is an unsupervised learning algorithm; it is the one that does not require labeled data inputs or any other type of feedback from the learning environment. In this paper: a thorough survey of the LLE algorithm, the implementation details, possible uses and extensions in assortment, and its relation to other eigenvector methods used for clustering and dimensionality reduction in nonlinear manner have been provided. Unsupervised algorithms attempt to solve any particular task in harder way than is necessary or, even the wrong problem altogether is an often made criticism. In the researcher's view, LLE removes much of the force behind this argument because it belongs to a new class of unsupervised learning algorithms. These new algorithms are distinguishable by global optimizations, simple functions of cost, and exhibit highly nonlinear behavior because of the factor that they do not make strong parametric assumptions. Broad usage of these algorithms is expected in numerous fields of information processing, and particularly the need to serve as a tool to simplify and accelerate other machine learning techniques that belongs to high dimensional spaces.

2.11) Unsupervised Learning (Z. Ghahramani)

The research paper provides a tutorial and overview of the field of unsupervised learning in perspective of statistical modeling. Unsupervised learning techniques can be motivated from; information theoretic and Bayesian principles. Basic models in unsupervised learning; including factor analysis, state-space models, some mixtures of Gaussian, hidden Markov models, ICA, PCA, and many variants plus extensions are briefly reviewed here. The researchers derive the EM algorithm and give an overview of basic concepts as graphical models, and inference algorithms as graphs; followed by a quick tour of approximate Bayesian inference, including Markov chain Monte Carlo (MCMC), Laplace approximation, variation approximations, and expectation propagation (EP). The aim of this research is to provide a high-level view of this field[11].

The formal study of learning systems is deduced from Machine learning; which is a field of research. It has found to be highly interdisciplinary field which acquires and constructs upon ideas from statistics, computer science (engineering), optimization theory, and numerous other disciplines of science and mathematics. This paper introduces in an adequately compact manner the key ideas prime to the sub-field of machine learning; known as unsupervised learning.

2.12) Building High-level Features Using Large Scale Unsupervised Learning(Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng)

The aim of this work is the use of unsupervised machine learning to build high-level, class specific feature detectors from unlabeled images and data sets. This approach is influenced by the neuro-scientific conjecture: there exist highly class-specific neurons in the human brain, informally known as "grandmother neurons". The role of labeled data to obtain these class-specific feature detectors is typically emphasized by Contemporary computer vision methodology [12]. For example; one needs a large collection of images labeled as containing faces, to build a face detector, usually with a boundary of dotted box around the face. Where labeled data are rare, a significant challenge for problems is posed and the need for large labeled sets is produced. Approaches that have not been shown to work well for building high-level features, make use of inexpensive unlabeled data although they are often preferred.This work inspects the achievability of building high level features: from only unlabeled data, a positive answer to it will give rise to two significant results. It is an inexpensive way to develop features from unlabeled data, more importantly; it answers an appealing question as to whether the particularity of the "grandmother neuron" could possibly be experienced from unlabeled data. Unsupervised feature learning and deep learning have surfaced as methodologies in machine learning from unlabeled data.

3. CONCLUSION

Research in the area of feature subset selection: unsupervised learning is quite young. Even though this paper has addressed some issues, more questions that need to be solved arise. Hartigan (1985) describes that no single criteria is best for all problems. A new and interesting angle is hierarchical clustering for feature selection Gennari, 1991; Fisher, 1996; Devaney and Ram, 1997; Talavera, 1999; Vaithyanathan and Dom, 2000, as hierarchical clustering provides groupings of datasets at various levels. Unsupervised learning techniques can be observed from the perspective of statistical modeling. For learning by using data and for interpretation under uncertainty, a consistent framework is provided via statistics. Many statistical models used for unsupervised learning can be made projects as in latent variable models and graphical models. Unsupervised learning systems for a range of different kinds of data: is illustrated by these models. Thus,

targeting on ways of approximating high dimensional sums and integrals has been followed by us from the study of unsupervised learning.

4. ANALYSIS DESCRIPTION

Results of analysis of evaluation parameters are shown in the table 3. Twelve techniques are discussed against 16 parameters. Important machine learning parameters are discussed in the first half and rest of them is discussed in the later. In Table 1 all the parameters along with their meaning and possible values are presented while table 2 shows the existence of these parameters in respective search paper techniques and related with quality. Necessary parameters such as Error rate, Precision, Accuracy, Recall, ROC (Receiver operating characteristics) factor, Case study for unsupervised machine learning techniques

Cost effective, performance, efficiency, reliability, maintainability, scalability, effortless and case study are discussed. Almost all mentioned techniques are cost effective and have good performance abilities. Performing activities/tasks without any failure is reliability in table 2.Analysis of reliability parameter shows that except K.Seymore all authors discuss reliability factor in research like Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng in table 3.New features can be easily adapted and are operable by the existing system or techniques is robustness shows in table 2.In Analysis table description there is no robustness factor in R. Raina, A.Madhava, and A. Y. Ng research [2]. Complexity is basically to what extent modules are inter-related and dependent in table 2.complexity is major factor in unsupervised machine learning algorithms so it is important L. Bo, X.Ren, and D. Fox[3] provide the techniques which is not fullfil the requirements for remove the complexity in table 3. Performance, Efficiency, Integrity. Reusability, Security are the other parameters describe in table 2 and analysis is shown in table 3. Case study is the basically reference to some experimentation in table 1. T.HOFMANN, J. G. Dy and C. E. Brodley, L. K. Saul and S. T. Roweis, Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng research not provide the information of parameter "case study" which is analyze by table 3 evaluation.Recall in table 1 is the important parameter for machine learning techniques which means that "Proportion of number of items as positive to total".

A.McCallum,K.Nigam,J.Rennie give brief discussion on recall on "Unsupervised Feature Learning for RGB-D Based Object Recognition " research paper[4].In table 2 Roc is the Organizing classifier based on their performance and there selecting and Graph visualization. Z. Ghahramani discuss about the ROC factor and error rate and precision in research paper "Unsupervised Learning"[11]. Datasets are unbalanced in information retrieval is the precision in table 1. Accuracy of a product show that there is number of correct predictions. Alcala-Fdez , L. Sanchez , S. Garcia , M. J. del Jesus discuss on accuracy in research paper "KEEL: a software tool to assess evolutionary algorithms for data mining problems"[7]. Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng also discuss on accuracy in research shows in table 3.

5. REFERENCES

[1] A. Jain. (2010). "Pattern Recognition Letters".2009 Elsevier B.V. All rights reserved. [On- line].31.(9),pp. 651–666. Available: http://www.sciencedirect.com/science/journal/01678655 [Sep.9, 2009].

[2] R. Raina, A.Madhava, and A. Y. Ng. "Large-scale Deep Unsupervised Learning using Graphics Processors." Internet: videolectures.net/site/normal_dl/tag=48368/ icml09_raina_lsd _01.pdf, 2009.

[3] L. Bo, X.Ren, and D. Fox. "Unsupervised Feature Learning for RGB-D Base Object Recognition." Internet: research.cs.washington.edu/istc/lfb/paper/iser12.pdf.

[4] A.McCallum,K.Nigam,and J.Rennie . "Building Domain Specfic Search Engines with Machine Learning Techniques." Internet: www.aaai.org/Papers/Symposia/ Spring/1999/SS-99.../SS99-03-006.pdf

[5] K.Seymore. "A Machine Learning Approach to Building Domain-Speci_C Search Engines." Www.Cora.Justresearch.Com. Available: qwone.com/~jason/papers/cora-ijcai99.pdf

[6] J.Magidson, J.K. Vermunt. "Latent class models for clustering: A comparison with K-means." In Canadian Journal of Marketing Research, Volume 20, 2002. Available: statisticalinnovations.com/articles/cjmr.pdf

[7] J. Alcalá-Fdez • L. Sánchez • S. García • M. J. del Jesus. "2008 KEEL: a software tool to assess evolutionary algorithms for data mining problems." Published online: 22 May Springer-Verlag 2008. Available: www.salleurl.edu/GRSI/docs/keel_softcomputing.pdf

[8] T.HOFMANN. "2001 Unsupervised Learning by Probabilistic Latent Semantic Analysis." Kluwer Academic Publishers. Manufactured in The Netherlands, Machine Learning, 42, 177–196, 2001. Available:www.cs.odu.edu/~sji/classes/dm-2013s/papers/pLSI.pdf

[9] J. G. Dy and C. E. Brodley. "Feature Selection for Unsupervised Learning." Journal of Machine Learning Research 5 (2004), 845–889. Available: http://www.jmlr.org/papers/volume5/dy04a/dy04a.pdf

[10] L. K. Saul and S. T. Roweis. "Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds." Journal of Machine Learning Research 4 (2003), 119-155. Available:http://www.jmlr.org/papers/volume4/saul03a/saul0 3a.pdf

[11] Z. Ghahramani. "Unsupervised Learning." Bousquet, O., Raetsch, G. and von Luxburg, U. (eds) Advanced Lectures on Machine Learning LNAI 3176. c Springer-Verlag. September 16,2004.Available:http://mlg.eng.cam.ac.uk/zoubin/pape rs/ul.pdf

[12] Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng. "Building High-level Features Using Large Scale Unsupervised Learning." 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012. Available: http://static.googleusercontent.com/media/research.googl e.com/en//archive/unsupervised_icml2012.pdf

TABLE 1   MACHINE LEARNING PARAMETERS, MEANINGS POSSIBLE VALUES

Serial #	Machine learning Parameters	Meanings	Possible values
1	*Error rate*	Against total predictions there is a number of incorrect predictions	Y, N, ND
2	*Precision*	Datasets are unbalanced in information retrieval	Y, N, ND
3	*Accuracy*	From total number of predictions there is number of correct	Y, N, ND
		predictions	
4	*Recall*	these are Proportion of number of items as positive to total	Y, N, ND
5	*ROC (Receiver operating*	Organizing classifier based on their performance and there	Y, N, ND
	characteristics) factor	selecting and Graph visualization	
6	*Case study*	These are the basically reference to some experimentation	Y, N, ND
7	*F1- score*	This parameter combines precision and recall into a single parameter with equal importance.	Y, N, ND

TABLE 2   EVALUATION CRITERIA FOR MODEL BASED SOFTWARE ENGINEERING FOR AUTOMATED TESTING

Sr	Quality Parameters	Meanings	Possible Values
1	*Reliability*	Performing activities/tasks without any failure.	Y, N, ND
2	*Efficiency*	Usage of minimum resources to produce high throughput.	Y, N, ND
3	*Cost effective*	Relation of expense to the productive process.	Y, N, ND
4	*Performance*	Recovery rate of the system under faulty inputs.	Y, N, ND
5	*Integrity*	Prevention against unauthorized access, protecting software and programs.	Y, N, ND
6	*Reusability*	The modules and techniques of previous system can be reused in a updated system.	Y, N, ND
7	*Robustness*	It is complementary to correctness. Accurate performance of a technique under cases not specified by the Requirements.	Y, N, ND
8	*Extendibility*	New features can be easily adapted and are operable by the existing system or techniques.	Y, N, ND
9	*Complexity*	To what extent modules are inter-related and dependent.	Y, N, ND
10	*Security*	The personal information of a specific user is not accessible to anyone.	Y, N, ND

TABLE 3 ANALYSIS OF EXISTING TECHNIQUES FOR UNSUPERVISED MACHINE LEARNING

#Sr		Precision	RateError	Recall	Accuracy	FactorROC	score-1F	studyCase	Integrity	Reliability	Reusability	Efficiency	Complexity	EffectiveCost	Robustness	Performance
1	A. Jain	Y	Y	Y	N	Y	Y	N	Y	ND	Y	Y	Y	Y	Y	N
2	R. Raina, A.Madhava, and A. Y. Ng	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
3	L. Bo, X.Ren, and D. Fox	N	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	ND	Y	N
4	A.McCallum,K.Nigam,J. Rennie	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	ND
5	K.Seymore	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	Y
6	J.Magidson, J.K. Vermunt	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y
7	Alcala-Fdez , L. Sanchez , S. Garcia , M. J. del Jesus	Y	ND	Y	Y	Y	Y	ND	ND	Y	Y	Y	Y	Y	N	N
8	T.HOFMANN	ND	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
9	J. G. Dy and C. E. Brodley	ND	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y
10	L. K. Saul and S. T. Roweis	N	Y	Y	Y	Y	Y	Y	N	N	N	ND	Y	Y	Y	Y
11	Z. Ghahramani	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N
12	Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng	Y	Y	Y	N	Y	N	Y	Y	N	Y	Y	ND	Y	Y	N