Approximate Bayesian Algorithm for Tensor Robust Principal Component Analysis

Andrej Srakar

Abstract   Recently proposed Tensor Robust Principal Component Analysis (TRPCA) (Lu et al. in Tensor robust principal component analysis: exact recovery of corrupted low-rank tensors via convex optimization, 2019 [14]) aims to exactly recover the low-rank and sparse components from their sum, extending the Low-Rank Tensor Completion model of Mu et al. (Lower bounds and improved relaxations for tensor recovery, 2013 [17]). We construct a Bayesian approximate inference algorithm for TRPCA, based on regression adjustment methods suggested in the literature to correct for high-dimensional nature of the problem (Blum in J Am Stat Assoc 105(491), 2010 [3]; Blum and François in Stat Comput 20(1):63–73, 2010 [4]). Our results are compared to previous studies using variational Bayes inference for tensor completion (Hawkins and Zhang in Conference: IEEE international conference on data mining, 2018 [11]). In a short application, we study spatiotemporal traffic data imputation using nine-week spatiotemporal traffic speed data set of Guangzhou, China.

Keywords   Tensor robust PCA · Low-rank · Tensor completion · Approximate bayesian computation · Regression adjustment · Variational bayes

AMS Subject Classification   62F15

1   Introduction: Tensor Robust Principal Component Analysis and Its Extensions

Classical Principal Component Analysis (PCA) is the most widely used statistical tool for data analysis and dimensionality reduction. It is computationally efficient and powerful for the data which are mildly corrupted by small noises. However, a major issue of PCA is that it is brittle to grossly corrupted observations or presence of outliers, which are ubiquitous in real world data. To date, a number of robust ver-

A. Srakar (✉)

Ljubljana and School of Economics and Business, Institute for Economic Research (IER), University of Ljubljana, Ljubljana, Slovenia
e-mail: andrej.srakar@ier.si

sions of PCA were proposed. But many of them suffer from the high computational cost. The recently proposed Robust PCA [7] is the first polynomial-time algorithm with strong performance guarantees. Suppose we are given a data matrix $X \in \mathbb{R}^{n_1 \times n_2}$ which can be decomposed as $X = L_0 + E_0$ where L_0 is low-rank and E_0 is sparse. It is shown in Candès et al. [7] that if the singular vectors of L_0 satisfy some incoherent conditions, L_0 is low-rank and E_0 is sufficiently sparse, then L_0 and E_0 can be recovered with high probability by solving the following convex optimization problem:

$$\min_{L,E} \|L\|_* + \lambda \|E\|_1, \quad s.t. \ X = L + E \tag{1}$$

where $\|L\|_*$ denotes the nuclear norm (sum of the singular values of L), $\|E\|_1$ denotes the ℓ1-norm (sum of the absolute values of all the entries in E) and

$$\lambda = 1/\sqrt{\max(n_1, n_2)} \tag{2}$$

To use RPCA, one has to first restructure/transform the multi-way data into a matrix. Such a preprocessing usually leads to the information loss and would cause performance degradation. To alleviate this issue, a common approach is to manipulate the tensor data by taking the advantage of its multi-dimensional structure. In this work, we study the Tensor Robust Principal Component (TRPCA) which aims to exactly recover a low-rank tensor corrupted by sparse errors.

Tensors are mathematical objects that can be used to describe physical properties, just like scalars and vectors. They are a generalisation of scalars and vectors; a scalar is a zero rank tensor, and a vector is a first rank tensor. The rank (or order) of a tensor is defined by the number of directions (i.e. dimensionality of the array) required to describe it.

The tensor multi rank of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector $r \in \mathbb{R}^{n_3}$ with its i-th entry as the rank of the i-th frontal slice of $\overline{\mathcal{A}}$, i.e., $r_i = rank(\overline{A}^{(i)})$. The tensor tubal rank, denoted as $\text{rank}_t(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ is from the t-SVD of $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$. That is

$$\text{rank}_t(\mathcal{A}) = \#\{i : \ \mathcal{S}(i, i, :) \neq 0\} = \max_i r_i \tag{3}$$

The tensor nuclear norm of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{A}\|_*$, is defined as the average of the nuclear norm of all the frontal slices of $\overline{\mathcal{A}}$, i.e., $\|\mathcal{A}\|_* = \frac{1}{n_3} \sum_{i=1}^{n_3} \left\|\overline{A}^{(i)}\right\|_*$.

Tensor Robust PCA (TRPCA) [14] aims to exactly recover a low-rank tensor corrupted by sparse errors. It aims to recover the low tubal rank component $\mathcal{L}_0$ and sparse component $\mathcal{E}_0$ from $\mathcal{X} = \mathcal{L}_0 + \mathcal{E}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ by convex optimization

$$\min_{\mathcal{L},\mathcal{E}} \|\mathcal{L}\|_* + \lambda \|\mathcal{E}\|_1, \quad s.t. \mathcal{X} = \mathcal{L} + \mathcal{E} \tag{4}$$

We firstly define few necessary concepts. An *orthogonal tensor* is a tensor $Q \in \mathbb{R}^{n \times n \times n_3}$ if it satisfies:

$$Q^* * Q = Q * Q^* = I \qquad (5)$$

where I is the identity tensor.

f-diagonal tensor is a tensor if each of its frontal slices is a diagonal matrix.

The Tensor Singular Value Decomposition (T-SVD) for third order tensors was proposed by Kilmer and Martin [13] and has been applied successfully in many fields, such as computed tomography, facial recognition, and video completion. Kilmer and Martin presented the concept of a tensor-tensor product with suitable algebraic structure such that classical matrix-like factorizations are possible. In particular, they gave the definition of the Tensor SVD (T-SVD) over this new product, and showed that truncating that expansion does give a compressed result that is optimal in the Frobenius norm.

Theorem 1 (Tensor Singular Value Decomposition (T-SVD) [13, 14]) *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then it can be factored as:*

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^* \qquad (6)$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor.

Alternative tensor factorization is CANDECOMP/PARAFAC (CP) and expresses a N-way tensor $\mathcal{A}$ as the sum of multiple rank-1 tensors:

$$\mathcal{A} = \sum_{r=1}^{R} s_r a_r^{(1)\circ}...^{\circ}a_r^{(N)}, \quad with \; a_r^{(k)} \in \mathbb{R}^{I_k} \qquad (7)$$

Our Bayesian approach is based on likelihood representation of the problem in (4) following variational Bayes perspective of Hawkins and Zhang [11]. Variational perspectives have been earlier adopted as solutions to intractable likelihood problems in matrix and tensor completion [1, 21, 22]. In general, likelihood free perspective is applied to matrix and tensor completion problems as computational complexity is significantly higher for high-dimensional data than that of other methods, and convergence is generally hard to assess [1, 5]. In order to address problems of high-dimensionality in approximate Bayesian inference regression adjustment is often recommended [2–5, 18] and we use it also in our analysis.

We assume that each tensor slice can be fit by $\mathcal{X}_k = \widetilde{\mathcal{X}}_k + \mathcal{S}_k + \mathcal{E}_k$, where $\widetilde{\mathcal{X}}_k$ is low-rank, $\mathcal{S}_k$ contains sparse outliers and $\mathcal{E}_k$ denotes dense noise with small magnitudes. We will denote with $\mathcal{Y}_{\Omega,k}$ the observation of current slice and by $\mathcal{S}_{\Omega,k}$ its outliers. For the likelihood function representation let τ specify the noise precision, $\widehat{a}_{i_n}^{(n)}$ the i_n-th row of $A^{(n)}$, λ controls the rank of factorization and $\{\gamma_{i_1,...,i_N}\}$ controls the sparsity of $\mathcal{S}_\Omega$.

We define the likelihood function and used priors for the transformed problem in (4) using Gaussian and Gamma priors as:

$$\left(\mathcal{Y}_\Omega \left| \{A^{(n)}\}_{n=1}^{N+1}, \mathcal{S}_\Omega, \tau \right.\right) = \prod_{(i_1,\ldots,i_n)\in\Omega} \mathcal{N}\left(\mathcal{Y}_{i_1\ldots i_N} \left| \left\langle \widehat{a}_{i_1}^{(1)}, \ldots, \widehat{a}_{i_N}^{(N)} \right\rangle + \mathcal{S}_{i_1\ldots i_N}, \tau^{-1} \right.\right) \tag{8}$$

$$(\Theta \mid \mathcal{Y}_\Omega)$$
$$= \frac{p\left(\mathcal{Y}_\Omega \left| \{A^{(n)}\}_{n=1}^{N+1}, \mathcal{S}_\Omega, \tau \right.\right) \left\{\prod_{n=1}^{(N+1)} p\left(A^{(n)}|\lambda\right)\right\} p\left(\lambda\right) p\left(\mathcal{S}_\Omega|\gamma\right) p\left(\gamma\right) p(\tau)}{p(\mathcal{Y}_\Omega)} \tag{9}$$

$$p\left(A^{(n)} \mid \lambda\right) = \prod_{i_n=1}^{I_n} \mathcal{N}\left(\widehat{a}_{i_n}^{(n)} \mid 0, \Lambda^{-1}\right), \quad \forall n \in [1, N+1] \tag{10}$$

$$p\left(\mathcal{S}_\Omega \mid \gamma\right) = \prod_{(i_1,\ldots,i_N)\in\Omega} \mathcal{N}\left(\mathcal{S}_{i_1\ldots i_N} \mid 0, \gamma_{i_1\ldots i_N}^{-1}\right) \tag{11}$$

$$p\left(\tau\right) = Ga\left(\tau | a_0^\tau, b_0^\tau\right), \quad p\left(\lambda\right) = \prod_{r=1}^{R} Ga\left(\lambda_r | c_0, d_0\right) \tag{12}$$

$$p(\gamma) = \prod_{(i_1,\ldots,i_N)\in\Omega} Ga\left(\gamma_{i_1\ldots i_N} \mid a_0^\gamma, b_0^\gamma\right) \tag{13}$$

2 Scheme of the Approximate Bayesian Algorithm

Modern statistical applications increasingly require the fitting of complex statistical models. Often these models are "intractable" in the sense that it is impossible to evaluate the likelihood function. This prohibits standard implementation of likelihood-based methods, such as maximum likelihood estimation or a Bayesian analysis. To overcome this problem there has been substantial interest in "likelihood-free" or simulation-based methods. Examples of such likelihood-free methods include simulated methods of moments [10], indirect inference (Gourièroux and Ronchetti 1993) [12], synthetic likelihood [9] and approximate Bayesian computation [19]. Of these, approximate Bayesian computation (ABC) methods are arguably the most common methods for performing Bayesian inference [15, 19]. For a number of years, ABC

methods have been popular in population genetics (e.g. Cornuet et al. [8]) and systems biology (e.g. Toni et al. [20]); more recently they have seen increased use in other application areas, such as econometrics [6] and epidemiology [9].

In our ABC algorithm for TRPCA we amend the variational Bayes perspective of Hawkins and Zhang [11] who use it on a temporally defined problem. We use regression adjustment based ABC using as a summary statistic array of tensor first and second moment defined as k-statistics [16] and tensor tubal rank as defined above.

Scheme of the algorithm:

Step 1. Simulate $\theta^{(i)}, i = 1, \ldots, n$ according to the prior structure defined above.
Step 2. Simulate $s^{(i)} = \text{array}(\mathcal{A})^{(i)}$ using the generative model $p(s^{(i)}|\theta^{(i)})$.
Step 3. Associate with each pair $(\theta^{(i)}, s^{(i)})$ a weight $w^{(i)} \propto K_h(\left\| s^{(i)} - s_{\text{obs}} \right\|)$, where K_h is a kernel function and $\| \ \|$ the multidimensional Euclidean distance.
Step 4. Fit a regression model where the response is θ and the predictive variables are the summary statistics s. Use a regression model to adjust the $\theta^{(i)}$ in order to produce a weighted sample of adjusted values. We use heteroskedastic adjustment, following Blum (2017), as follows:

$$\theta_{c'}^{(i)} = \widehat{m}(s_{\text{obs}}) + \frac{\widehat{\sigma}(s_{\text{obs}})}{\widehat{\sigma}(s^{(i)})}(\theta^{(i)} - \widehat{m}(s^{(i)})) \tag{14}$$

where $\widehat{m}$ and $\widehat{\sigma}$ are the standard estimators of the conditional mean and of the conditional standard deviation.

3   Numerical Experiments and Application

With the development of intelligent transportation systems, large quantities of urban traffic data are collected on a continuous basis from various sources. These data sets capture the underlying states and dynamics of transportation networks and the whole system. In general, traffic data register full spatial and temporal features, together with some other site-specific attributes. Usually, we can organize the spatiotemporal traffic data into a multi-dimensional structure. Combined with information from other links in a city, the overall spatiotemporal data can be structured as a multi-dimensional array, which is often referred to as a tensor. A common drawback that undermines the use of such spatiotemporal data is the "missingness" problem, which may result from various factors such as hardware/software failure, network communication problems, and zero/limited reports from floating/crowdsourcing systems.

To demonstrate the performance of this model, in this section we conduct numerical experiments based on a large-scale traffic speed data set collected in Guangzhou, China. The data set is generated by a widely-used navigation app on smart phones. The data set contains travel speed observations from 214 road segments in two months (61 days from August 1, 2016 to September 30, 2016) at 10-min interval (144 time intervals in a day). The speed data can be organized as a third-order tensor (road

segment $\times$ day $\times$ time interval). Among the 1.88 million elements, about 1.29% are not observed or provided in the raw data.

In Tables 1 and 2 we compare performance of different models applied to several scenarios. We compare: Bayesian Gaussian CANDECOMP/PARAFAC (BGCP) tensor decomposition model, high accuracy low-rank tensor completion (HaLRTC) (Liu et al. 2013), which is used in Ran et al. (2016), SVD-combined tensor decomposition (STD) (Chen et al. 2018), DA (daily average) fills the missing value with an average of observed data (over different days) for the same road segment and the same time window (Li et al. 2013). kNN is another baseline method where the neighbors refer to road segments. Finally, TRPCA-VAR and TRPCA-ABC refer to tensor robust PCA specification in variational Bayes and approximate Bayesian computation algorithm form. The mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to evaluate model performance. Our first experiment examines the performance of different models and different representations in the random missing scenario. In the second experiment, we present a more realistic temporally correlated missing scenario. From the original data set we create five novel datasets with different missing rates ranging from 10 to 50%. We use two data representations: matrix representation (A) and third-order tensor representation (B).

As can be seen from the tables (the best models are marked in bold), for the random missing scenario, frequently the Variational Bayes specification performs best. On the other hand, our ABC approach performs very well in the second, temporally correlated missing scenario.

4 Conclusion

Our article provides an initial step in the development of ABC algorithms for tensor completion and tensor principal component analysis. We upgrade the tensor robust PCA approach of Lu and coauthors using approximate Bayesian perspective which provides ground for further research in the area of Bayesian approaches in matrix and tensor completion. Also, our article provides additional information on approximate Bayesian approaches to high-dimensional problems in statistics.

Few possible extensions of our work and pathways for future work seem apparent:

- Other possibilities of the ABC algorithms (such as SMC, HMC, other regression and marginal adjustment approaches) integrated nested Laplace approximation, including additional upgrades of the variational approach of Hawkins and Zhang should lead to more evidence on methodological possibilities to approach matrix and tensor completion from a Bayesian computational perspective.
- Different loss and divergence measures (for example Bregman type divergence measures) could be tested and asymptotics of the approach developed.
- Extension to different type of tensor measures and different specifications of the tensor robust PCA (the specification we use is only one of the possible ones) as well as extensions to any type and size of a tensor.

Table 1 The imputation performance of BGCP, HaLRTC, STD, DA (daily average), kNN, TRPCA-VAR and TRPCA-ABC for two data representations in the first scenario (best models are highlighted in bold)

		10%		20%		30%		40%	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
(A)	BGCP(50)	0.0937	3.9981	0.0952	4.0467	0.0962	4.0903	0.0976	4.1457
	BGCP(80)	0.0925	3.9483	0.0941	3.9958	0.0951	4.0449	0.0968	4.1091
	BGCP(110)	0.0920	3.9303	0.0937	3.9790	0.0948	4.0292	0.0965	4.0937
	HaLRTC	0.0957	3.9666	0.0976	4.0232	0.0991	4.0820	0.1009	4.1467
	DA	0.1213	5.1778	0.1218	5.1905	0.1217	5.1977	0.1217	5.1993
	kNN(10)	0.1303	5.1101	0.1314	5.1486	0.1322	5.1966	0.1333	5.2565
	TRPCA-VAR	0.1003	4.3452	0.0786	4.4013	0.1045	4.4291	0.0948	4.4778
	TRPCA-ABC	0.1083	4.3652	0.1026	4.4063	0.0825	4.4141	0.1008	4.4854
(B)	BGCP(50)	0.0862	3.7097	0.0867	3.7199	0.0867	3.7298	0.0867	3.7317
	BGCP(80)	0.0823	3.5614	0.0827	3.5660	0.0827	3.5775	0.0829	3.5851
	BGCP(110)	0.0795	3.4521	0.0798	3.4531	0.0799	3.4655	**0.0801**	**3.4756**
	HaLRTC	0.0777	3.1917	0.0815	3.3324	0.0850	3.4748	0.0887	3.6143
	STD	0.0888	3.7708	0.0911	3.8308	0.0936	3.9286	0.0963	4.0265
	TRPCA-VAR	**0.0639**	**3.5291**	**0.0804**	**3.5534**	**0.0666**	**3.6102**	0.0929	3.6656
	TRPCA-ABC	0.0939	3.5441	0.0964	3.5924	0.0856	3.6332	0.0979	3.6886

Table 2 The imputation performance of BGCP, HaLRTC, STD, DA, kNN, TRPCA-VAR and TRPCA-ABC for two data representations in the second scenario (best models are highlighted in bold)

		10%		20%		30%		40%	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
(A)	BGCP(15)	0.1011	4.2458	0.1013	4.2674	0.1020	4.3162	0.1031	4.3915
	BGCP(20)	0.1005	4.2307	0.1010	4.2755	0.1017	4.3229	0.1031	4.4124
	HaLRTC	0.1015	**4.1322**	0.1022	4.1716	0.1035	4.2372	0.1057	4.3232
	DA	0.1208	5.1128	0.1207	5.1353	0.1200	5.1408	0.1196	5.1434
	kNN(13)	0.1342	5.1714	0.1340	5.1983	0.1346	5.2591	0.1388	5.4405
	TRPCA-VAR	0.0906	4.5926	0.1158	4.5866	0.1204	4.6352	0.1291	4.7332
	TRPCA-ABC	0.1226	4.5736	0.0978	4.5866	0.1264	4.6362	0.1001	4.7352
(B)	BGCP(15)	0.0992	4.1760	0.0995	4.1949	0.0999	4.2425	0.1001	**4.2881**
	BGCP(20)	0.0980	4.1413	0.0984	**4.1477**	**0.0980**	**4.1857**	0.1006	4.4556
	HaLRTC	0.1033	4.1576	0.1046	4.2086	0.1062	4.2792	0.1088	4.3813
	STD	0.1019	4.1881	0.1054	4.3300	0.1068	4.4029	0.1115	4.5573
	TRPCA-VAR	0.0956	4.1598	0.1100	4.1903	0.1047	4.2626	0.0943	4.4116
	TRPCA-ABC	**0.0746**	4.1758	**0.0740**	4.2343	0.1167	4.2866	**0.0853**	4.4006

References

1. Babacan, S.D., Luessi, M., Molina, R., Katsaggelos, A.K.: Sparse Bayesian methods for low-rank matrix estimation. IEEE Trans. Signal Process **60**(8), 3964–3977 (2012)
2. Beaumont, M.A., Zhang, W., Balding, D.J.: Approximate Bayesian computation in population genetics. Genetics **162**, 2025–2035 (2002)
3. Blum, M.G.: Approximate Bayesian computation: a nonparametric perspective. J. Am. Statistical Assoc
4. Blum, M.G., François, O.: Non-linear regression models for approximate Bayesian computation. Stat. Comput. **20**(1), 63–73 (2010)
5. Blum, M.G.B., Nunes, M.A., Prangle, D., Sisson, S.A.: A comparative review of dimension reduction methods in approximate Bayesian computation. Stat. Sci. **28**, 189–208 (2013)
6. Calvet, L.E., Czellar, V.: Accurate methods for approximate Bayesian computation filtering. J. Financ. Econ. **13**(4), 798–838 (2015)
7. Candès, E.J., Li, X.D., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM **58**(3) (2011)
8. Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., Estoup, A.: Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics **24**(23), 2713–2719 (2008)
9. Drovandi, C.C., Pettitt, A.N.: Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics **67**(1), 225–233 (2011)
10. Duffie, D., Singleton, K.J.: Simulated moments estimation of Markov models of asset prices. Econometrica **61**(4), 929–952 (1993)
11. Hawkins, C., Zhang, Z.: Variational Bayesian inference for robust streaming tensor factorization and completion. In: Conference: IEEE International Conference on Data Mining, Nov 2018. Available as arXiv:1809.01265v1 (2018)
12. Heggland, K., Frigessi, A.: Estimating functions in indirect inference. J. Roy. Stat. Soc. Ser. B **66**, 447–462 (2004)
13. Kilmer, M.E., Martin, C.D.: Factorization strategies for third-order tensors. Linear Algebra Appl. **435**(3), 641–658 (2011)
14. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: Tensor robust principal component analysis: exact recovery of corrupted low-rank tensors via convex optimization. Available as arXiv:1708.04181v3 (2019)
15. Martin, G., Frazier, D., Robert, C.P.: Computing Bayes: Bayesian computation from 1763 to the 21st century. Available as arXiv:2004.06425 (2020)
16. McCullagh, P.: Tensor Methods in Statistics, 2nd edn. Dover Books on Mathematics (2018)
17. Mu, C., Huang, B., Wright, J., Goldfarb, D.: Square deal: lower bounds and improved relaxations for tensor recovery. Available as arXiv:1307.5870v2 (2013)
18. Nott, D.J., Ong, V.M.-H., Fan, Y., Sisson, S.A.: High-dimensional ABC. In: Handbook of Approximate Bayesian Computation, pp. 211–242 (2018)
19. Robert, C.P.: Approximate Bayesian computation: a survey on recent methods. In: Cools, R., Nuyens, D. (eds.) Monte Carlo and Quasi-Monte Carlo Methods (MCqMC), pp. 195–205. Springer, Berlin (2014)
20. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. Roy. Soc. Interface **6**(31), 187–202 (2009)
21. Zhao, Q., Zhang, L., Cichocki, A.: Bayesian CP factorization of incomplete tensors with automatic rank determination. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1751–1763 (2015)
22. Zhao, Q., Zhou, G., Zhang, L., Cichocki, A., Amari, S.-I.: Bayesian robust tensor factorization for incomplete multiway data. IEEE Trans. Neural Networks Learn. Syst. **27**(4), 736–748 (2016)