Deep learning based Feature Selection Study

Sanghyuck Lee

2021. 08 31.

Outline

- Dynamic Instance-wise Joint Feature Selection and Classification
- 중견 연구 과제
- 갑상선 연구 과제
- 산학 협력 과제
- 11월 특강 자료 작성

Papers for Instance-Wise Feature Selection

Yoon, Jinsung, James Jordon, and Mihaela van der Schaar. "INVASE: Instance-wise variable selection using neural networks." *International Conference on Learning Representations*. 2018.

=> ICLR

Tonekaboni, Sana, et al. "What went wrong and when? Instance-wise feature importance for time-series black-box models." *Advances in Neural Information Processing Systems* 33 (2020).

=> NeurIPS

Panda, Pranoy, Sai Srinivas Kancheti, and Vineeth N. Balasubramanian. "Instance-wise Causal Feature Selection for Model Interpretation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

=> CVPR

Liyanage, Yasitha Warahena, and Daphney–Stavroula Zois. "Optimum Feature Ordering for Dynamic Instance–Wise Joint Feature Selection and Classification." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

=> ICASSP

Paper Review

Paper Title: **Optimum Feature Ordering for Dynamic Instance-wise Joint Feature Selection and Classification [1]**

Authors: Yasitha Warahena Liyanage, Daphney-Stavroula Zois and Charalampos Chelmis

Conference: ICASSP 2021

Abstract of the Paper

We introduce a supervised machine learning framework to perform **joint feature selection and classification** individually for each data instance **during testing**. In contrast to our prior work, we decide both **the order and the number of features for each data instance**. Specifically, our proposed solution dynamically selects the feature to review at each stage based on the already observed features and **stops the selection process to make a prediction once it determines no classification improvement can be achieved.** To gain insights, we analyze the properties of the proposed solution. Based on these properties, we propose a fast algorithm and demonstrate its effectiveness compared to the state–of–the–art using 4 publicly available datasets.

INTRODUCTION

In many real world applications (e.g., medical diagnosis, disaster prediction), **features are not freely available to acquire**, while **accurate, time–sensitive and interpretable decisions** are needed. For instance, consider the case where a doctor aims to diagnose a patient (classification decision) as quickly as possible by conducting the minimum number of tests (features) due to both the cost of the tests and the time sensitivity of the decision. Note that a **different set of tests** may be appropriate **for each individual patient** (data instance). For example, relevant features for predicting heart failure may differ across patient subgroups [1]. At the same time, **the order** by which test are conducted for each patient does not only affect the cost, but **also the classification decision** [2].

INTRODUCTION

In our prior work [3, 4], we studied the problem of instance–wise dynamic feature selection and classification for independent/correlated features. The goal was to determine the number of features needed to classify an instance and the classification strategy **when the order by which the features are reviewed is fixed and common for all instances.** Herein, we remove this restrictive assumption and consider the more general problem of determining also the order by which features must be reviewed. Thus, **we derive the optimum feature ordering**, the number of features and the classification strategy that needs to be adopted for each data instance individually when features sequentially arrive one at a time during testing. We also analyze the theoretical properties of this optimum solution. Specifically, we show that the corresponding functions are continuous, concave, and piece–wise linear on the domain of a sufficient statistic, and use these properties to derive an efficient implementation. In our experiments, we observe that **both the order** by which our framework reviews features and **the number of features** used for each test instance **are varying**. In addition, **less number of features on average** is needed to **achieve comparable performance** with the state–of–the-art.

PROBLEM DESCRIPTION

Our goal is to assign each data instance s to one out of L possible classes. At the same time, we wish to select the **most informative features** to use to reach such a classification decision, under the constraint that features **arrive sequentially one at a time** during testing.

In this context, at each time, we have to decide between: i) **stopping** and optimally **classifying** the **current** data **instance** based on the **reviewed features,** or ii) **continue and selecting the next feature** from the remaining set of available features.

PROBLEM DESCRIPTION

$\sigma, \sigma(R)$: Feature Selection Strategy - the order by which features are selected and the feature at which the sequential selection process stops, respectively.

For example, if $K = 3$, the $\sigma = (F_3, F_1, F_2)$ denotes a valid feature ordering, while $\sigma(R = 2) = F_1$ indicates that our framework stops after acquiring the second feature F_1 in the ordering σ.

$D_{\sigma(R)}$: Classification Strategy

For example, if $K = 3$, $L = 2$ and $\sigma = (F_3, F_1, F_2)$, then the event $\{D_{\sigma(R=2)} = 1\}$ represents deciding in favor of class c_1 based on the feature assignments $\{f_3, f_1\}$.

PROBLEM DESCRIPTION

In order to find the optimum ordering σ^* , the optimum stopping feature $\sigma^*(R^*)$ and the optimum classification strategy $D^*_{\sigma^*(R^*)}$ that can lead to an accurate classification decision for each data instance $s \in S$, we propose to solve the following optimization function:

$$\min_{\sigma,\sigma(R),D_{\sigma(R)}} J(\sigma,\sigma(R),D_{\sigma(R)})$$

where $J\big(\sigma,\sigma(R),D_{\sigma(R)}\big) = \mathbb{E}\big\{\Sigma_{k=1}^{R} e\big(F_{\sigma(k)}\big)\big\} + \Sigma_{j=1}^{L}\Sigma_{i=1}^{L} Q_{ij} P\big(D_{\sigma(R)} = j, C = c_i\big)$.

The first term represents the cost of reviewing features, and the second term penalizes the misclassification cost of decision $D_{\sigma(R)}$.

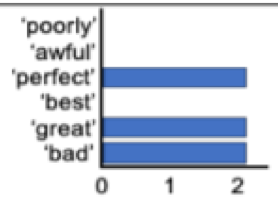
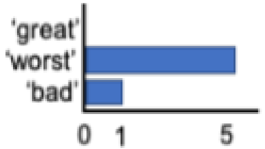
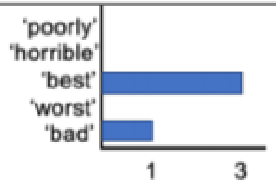
Experiments of the Paper

Table 1. Performance comparison with baselines. "Acc", and "Feat" stands for accuracy, and the average number of features per data instance, respectively. "All" represents using all features.

Method	MLL		Spambase		Lung2		Car	
	Acc	Feat	Acc	Feat	Acc	Feat	Acc	Feat
IFCO	**1.00**	**3.20**	**0.813**	**3.01**	**0.887**	**3.94**	**0.857**	**8.64**
MB [4]	**1.00**	**4.88**	0.741	3.08	0.842	3.96	0.539	5.63
ASSESS [3]	**1.00**	**5.07**	0.847	7.47	0.882	15.6	0.810	12.91
OFS-Density [7]	0.960	11.0	0.787	7.60	**0.912**	**16.2**	0.597	6.80
SAOLA [8]	0.867	28.0	0.824	24.6	0.882	28.2	0.798	41.4
OSFS [9]	0.800	3.00	0.801	33.8	0.847	5.80	0.556	5.20
FAST–OSFS [9]	0.800	5.00	0.801	33.8	0.842	9.40	0.608	8.40
Lasso	**1.00**	**4.00**	0.902	29.6	0.685	9.40	0.551	28.8
Tree [5]	0.933	100	0.947	18.2	0.897	207	0.752	429
PCA	0.667	36.0	0.693	1.00	0.897	88.4	0.391	91.0
SVM–G	**1.00**	**All**	0.834	All	0.788	All	0.563	All
R–Forest	**1.00**	**All**	0.940	All	0.911	All	0.758	All
XG–Boosting	0.733	All	**0.955**	**All**	0.906	All	**0.844**	**All**

Experiments of the Paper

Table 2. Words (features) picked by IFCO are highlighted in yellow. The true/predicted label is given at the end of each review. The second column reports features selected in ascending order (Y–axis) versus feature value (X–axis).

IMDB Review Text (**True Label, Predicted Label**)	
I had read up on the film . . . I wasn't expecting anything great, figured it would be mostly fluff but hopefully not a totally bad experience. I have to admit I was pleasantly surprised. The dialogue was pitch perfect, most of the actors were exceptionally good and it flowed nicely. Ash Christian was perfect, . . . Ashley Fink is gem, a great young character actress that hopefully will get more work. There are moments in the film that could have used some work, but all in all not a bad time at the cinema. . . . (**positive, positive**)	
This movie is the worst thing ever created by humans. You think manos is the worst movie ever? It doesn't even come close to this garbage. I dont even know where to begin. The "russian" commander and the rebel chic are the worst "actors" ever to appear in a movie. . . . the goofiest rape scene ever filmed, and the worst acting ever put on film. This movie deserves to be more well known among bad movie fans. Definitely the worst movie ever made. (**negative, negative**)	
Do-It-Yourself indie horror auteur Todd Sheets . . . and a trio of hottie sisters all have to do their best to survive this harrowing ordeal. . . . Let's not forget the ridiculous ending in which several of our survivors stumble across a few vials of flesh-eating bacteria to use on the shambling undead hordes. Sure, this flick is pure dreck, but it has a certain endearingly abominable quality to it that in turn makes it a great deal of so-awful-it's-awesome Grade Z fun for hardcore aficionados of bad fright fare. (**positive, negative**)	
Tommy Lee Jones was the best Woodroe and no one can play Woodroe F. Call better than he. Not only was he the first and best, he was the only person that could portray his grief and confusion. It was a bad let-down and I'm surprised I even made myself watch it. . . . The first movie was the best and the only . . . (**negative, positive**)	

Conclusion of the Paper

We presented an instance–wise dynamic joint feature selection and classification framework that selects both the order and the number of features for each data instance individually. We showed that the functions related to the optimum solution are continuous, concave and piece–wise linear on the domain of a sufficient statistic. Using these properties, we proposed IFCO, and validated its effectiveness compared to prior work using real–world datasets. In our future work, we plan to extend this work to regression settings.

Conclusion of Another Paper [2]

Liyanage, Yasitha Warahena, and Daphney–Stavroula Zois. "Optimum Feature Ordering for Dynamic Instance–Wise Joint Feature Selection and Classification." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

Liyanage, Yasitha Warahena, Daphney-Stavroula Zois, and Charalampos Chelmis. "Dynamic Instance-Wise Joint Feature Selection and Classification." *IEEE Transactions on Artificial*

The limitations listed above hint to corresponding **promising future research directions**. In addition, recent work has argued that different features for different instances can be used to **interpret machine learning outcomes** [6]–[8]. In this context, <u>the proposed methods can potentially **provide insights into the labeling of individual data instances based on evaluated features**</u>. A more thorough evaluation of the proposed methods will be carried out in the future on their efficacy for model interpretation, along with other necessary improvements toward this direction.

Discussion after Reviewing the Papers

Causal: Interpretability / X-Y(predicted)'s relevant -> Important Patches Selection

Dynamic/Joint: Joint FS and CLF-> Higher performance, Less number of features

FS based on DNNs or FS with DNNs (e.g. DeepPINK [3], Knockoffs studies, …)

=> Pruning (e.g. SCOP [4]) and Saliency map (e.g. DANCE [5])

⇒Interpretability of DNNs with FS (e.g. Instance-wise FS [6])

==> Interpretability of DNNs is One of major study in recent AI field.

ㄴMy interest

중견 연구 제안서 작성

- 가제:

  한글: 의료 분야내 주요 관심 구조물 세그멘테이션을 위한 사전훈련 최신 딥러닝 모델 개발

  영문: Development of Pre-trained Deep Learning Model for Segmentation of Structures of Major Interest in the Medical Field

- 연구 목표:

  본 과제에서는 몇몇 주요한 해부학적 구조에 대해서 세계 최고 수준의 성능을 갖는 사전학습된 세그멘테이션 모델을 배포함으로써, 다양한 의료분야 응용에 성공적으로 기여하는 것을 최종 목표로 하며, 이를 위해 다음과 같은 연차별 연구 목표를 설정하여 추진한다.

중견 연구 제안서 작성

1. 연구의 목표 (2페이지)

- 최종 연구 목표
- 연차별 목표 및 개요도
- 연차별 연구 개발 내용, 창의성 및 도전성, 추진 전략, 주요 결과물

2. 연구의 필요성 (2페이지)

- 4~5개로 나누어 총 2페이지

3. 연구자의 연구 수행역량 (1페이지)

- 연구실 스펙 및 노하우 등 자랑거리

중견 연구 제안서 작성

4. 연구의 추진 전략 및 방법 (연차별 1.5페이지)

- 연차별 상세 내용 및 주요 기술 설명 (그림 포함하여, 연차당 3가지 정도)

5. 연구결과의 중요성 (0.5페이지)

- 학문적/기술적 파급 효과

- 경제적/사회적 파급 효과

6. 연구기간 및 연구비 적정성 (0.5페이지)

- 추진일정 등

- 연구비 적정성

중견 연구 제안서 작성

연차	데이터셋 이름	촬영 방식	차원	촬영 범위	타겟	샘플수	포함 근거
1	LIDC IDRI	CT	3D	흉부, 폐	폐 결절	1010	국책사업에 언급된 한국인 호발암 3위, 코로나 시국 관련
2	CAU Orbit	CT	3D	머리, 안와 부근	안와내 구조물	1237	갑상선 안병증 치료 계획에 중요 역할
3	ACRIN 6664	CT	3D	결장 부근	결장	825	국책사업에 언급된 한국인 호발암 2위 (대장암)
4	CBIS-DDSM	MG	2D	유방	유방암	1566	국책사업에 언급된 한국인 호발암 5위
5	Prostate MRI US Biopsy	MR	3D	전립선 부근	전립선	1151	국책사업에 언급된 한국인 호발암 7위

갑상선 연구

- 완료된 세그멘테이션 자료 전달 완료
- 추가 연구원 자료 일부 전달 완료, 이력서 및 기타 교육 이수증 전달 예정
- Activity 실험 아직 미진행

산학 협력 프로젝트

- 특허 출원: 이번 주 중으로 명세서 초안 수신 예정

11월 특강 자료 작성

- 주제: '사회적 약자의 보편적 복지 향상을 위해 활용될 수 있는 AI 기술 조사'
- 하위 선별 주제:
  1) 독거노인 등 취약 계층을 위한 AI로봇, AI스피커 - 우울증, 치매예방 및 생활관리, 위급상황시 대처에 대한 내용
  2) 외국인 및 소년 소녀 가장을 위한 교육 인공지능
  3) 장애인을 위한 생활 보조 및 학습 보조 AI - 장애인의 삶의 질 향상
  4) 시각장애인을 위한 AI시스템
  5) AI를 이용한 아동학대 탐지

⇒ 여쭤볼 사항: '사회적 약자' 키워드와 '보편적 복지' 키워드

Ex) 무상급식, 반값등록금

보편적 복지	선별적 복지
보편적 복지를 주장하는 입장에서는 복지를 국민의 하나의 권리라고 말합니다. 즉, 소득에 상관없이 세금을 낸 사회구성원이라면 누구든 혜택을 누려야 한다는 것입니다.	선별적 복지를 주장하는 측에서는 효율적인 복지를 말합니다. 모두가 혜택을 받을 필요가 없고, 정말 국가의 도움이 없이는 생활할 수 없는 최저계층에게만 복지 서비스를 제공해야 한다는 것입니다.

Ex) 기초생활 보장제도

Reference

[1] Liyanage, Yasitha Warahena, and Daphney–Stavroula Zois. "Optimum Feature Ordering for Dynamic Instance–Wise Joint Feature Selection and Classification." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

[2] Liyanage, Yasitha Warahena, Daphney-Stavroula Zois, and Charalampos Chelmis. "Dynamic Instance-Wise Joint Feature Selection and Classification." *IEEE Transactions on Artificial Intelligence* (2021).

[3] Lu, Yang Young, et al. "DeepPINK: reproducible feature selection in deep neural networks." *arXiv preprint arXiv:1809.01185* (2018).

[4] Tang, Yehui, et al. "Scop: Scientific control for reliable neural network pruning." *arXiv preprint arXiv:2010.10732* (2020).

[5] Lu, Yang Young, et al. "DANCE: Enhancing saliency maps using decoys." *International Conference on Machine Learning*. PMLR, 2021.

[6] Yoon, Jinsung, James Jordon, and Mihaela van der Schaar. "INVASE: Instance-wise variable selection using neural networks." *International Conference on Learning Representations*. 2018.