



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Лаборатория алгоритмов и технологий
анализа сетевых структур

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ МУЛЬТИМЕДИЙНЫХ ДАННЫХ

текущее состояние и перспективы развития
в лаборатории ЛАТАС

Нижний Новгород, 2019



ВЫЧИСЛИТЕЛЬНО ЭФФЕКТИВНЫЕ АЛГОРИТМЫ РАСПОЗНАВАНИЯ ОБРАЗОВ

ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ (1)

Приближенный поиск ближайшего соседа

Метод максимального правдоподобного перебора

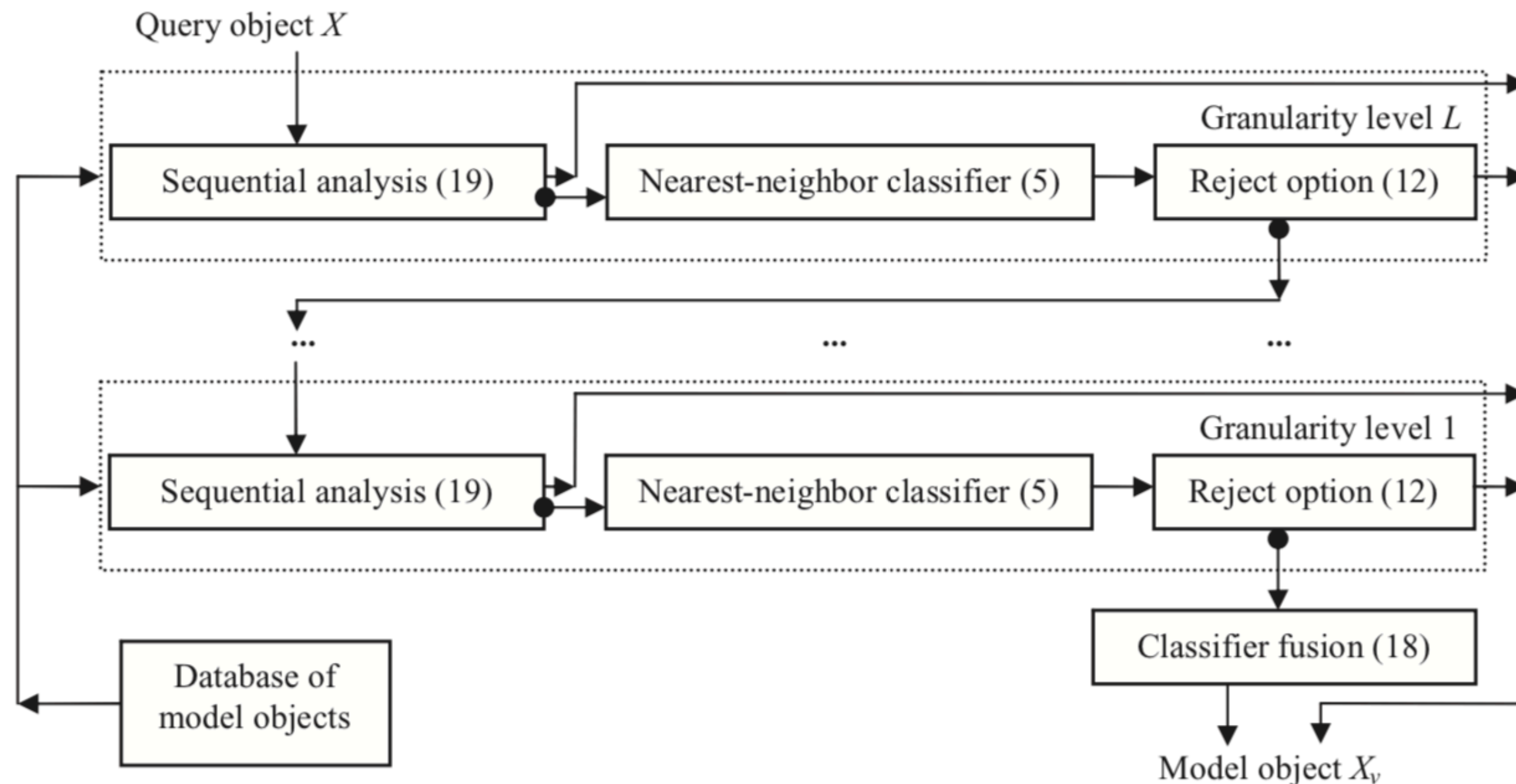
$$r_{k+1} = \underset{\nu \in \{1, \dots, R\} - \{r_1, \dots, r_k\}}{\operatorname{argmax}} \left(p_\nu \cdot \prod_{i=1}^k f(\rho(X, X_{r_i}) | W_\nu) \right),$$

$$r_{k+1} = \underset{\mu \in \{1, \dots, R\} - \{r_1, \dots, r_k\}}{\operatorname{argmin}} \left(\sum_{i=1}^k \varphi_\mu(r_i) - \ln p_\mu \right)$$
$$\varphi_\mu(r_i) \approx \frac{(\rho(X, X_{r_i}) - \rho_{\mu, r_i})^2}{\rho_{\mu, r_i}}$$

- Savchenko A.V. Maximum-Likelihood Approximate Nearest Neighbor Method in Real-time Image Recognition // Pattern Recognition. – 2017. – Vol. 61 . – pp. 459-469.
- Savchenko A.V. Clustering and maximum likelihood search for efficient statistical classification with medium-sized databases // Optimization Letters.– 2017.– Vol. 11(2).– P. 329-341
- Savchenko A.V. Deep neural networks and maximum likelihood search for approximate nearest neighbor in video-based image recognition, Optical Memory and Neural Networks (Information Optics), 2017, vol. 26, no. 2, pp. 129–136
- Savchenko A.V. Deep Convolutional Neural Networks and Maximum-Likelihood Principle in Approximate Nearest Neighbor Search. In: Alexandre L., Salvador Sánchez J., Rodrigues J. (eds) Iberian Conference on Pattern Recognition and Image Analysis. IbPRIA 2017. Lecture Notes in Computer Science, vol 10255, pp. 42-49. Springer, Cham (2017)
- Савченко А.В. Метод максимально правдоподобного перебора в задаче классификации кусочно-однородных объектов // Автоматика и телемеханика.– 2016.– № 3.– С. 99-108

ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ (2)

Метод последовательных тернарных решений



$$POS_{\rho_1^{(l)}} = \{ \rho_{KL}^{(l)}(X_1, X_2) | X_1, X_2 \in \mathbf{X}, \rho_{KL}^{(l)}(X_1, X_2) > \rho_1^{(l)} \}$$

$$NEG_{\rho_0^{(l)}} = \{ \rho_{KL}^{(l)}(X_1, X_2) | X_1, X_2 \in \mathbf{X}, \rho_{KL}^{(l)}(X_1, X_2) < \rho_0^{(l)} \}$$

$$BND_{\rho_0^{(l)}; \rho_1^{(l)}} = \{ \rho_{KL}^{(l)}(X_1, X_2) | X_1, X_2 \in \mathbf{X}, \rho_0^{(l)} \leq \rho_{KL}^{(l)}(X_1, X_2) \leq \rho_1^{(l)} \}$$

- Savchenko, A.V. Fast multi-class recognition of piecewise regular objects based on sequential three-way decisions and granular computing // Knowledge-Based Systems.– 2016.– Vol. 91.– P. 250–260
- Savchenko A.V. Sequential Three-Way Decisions in Efficient Classification of Piecewise Stationary Speech Signals. In: Polkowski L. et al. (eds) Rough Sets. IJCRS 2017. Lecture Notes in Computer Science/LNAI, vol 10314, pp. 264-277. Springer, Cham (2017)
- Savchenko A.V. Granular Computing and Sequential Analysis of Deep Embeddings in Fast Still-to-Video Face Recognition, Proceedings of IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI 2018), pp. 515-520, 2018



ПЕРСПЕКТИВЫ

Последовательный анализ и сверточные нейронные сети

1. Метод ближайшего соседа и векторы признаков (embeddings)

$$POS(c) = \left\{ \mathbf{x} \in \mathbf{X} \mid (c = c_1(\mathbf{x})) \& \left(\forall_{i:i \neq c} \frac{\rho_c(\mathbf{x})}{\rho_i(\mathbf{x})} < \delta_{1/*}(i) \right) \right\}$$
$$NEG(c) = \left\{ \mathbf{x} \in \mathbf{X} \mid \frac{\rho_{c_1(\mathbf{x})}(\mathbf{x})}{\rho_c(\mathbf{x})} \geq \delta_{1/*}(c) \right\}$$
$$BND(c) = \left\{ \mathbf{x} \in \mathbf{X} \mid \left(\frac{\rho_{c_1(\mathbf{x})}(\mathbf{x})}{\rho_c(\mathbf{x})} < \delta_{1/*}(c) \right) \& \left(\exists_{i:i \neq c} \frac{\rho_{c_1(\mathbf{x})}(\mathbf{x})}{\rho_i(\mathbf{x})} < \delta_{1/*}(i) \right) \right\}$$

2. Последовательная классификация (linear svm) выходов промежуточных слоев

3. Совместная работа с Соколовой А.

$$\rho(\tilde{\mathbf{x}}^{(l+1)}, \tilde{\mathbf{x}}_r^{(l+1)}) = \rho(\tilde{\mathbf{x}}^{(l)}, \tilde{\mathbf{x}}_r^{(l)}) + \sum_{d=(l-1)m+1}^{lm} \rho(\tilde{x}_d, \tilde{x}_{r;d})$$



РАСПОЗНАВАНИЕ ОБРАЗОВ ПРИ НАЛИЧИИ МАЛОГО ЧИСЛА ЭТАЛОНОВ



ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

Модификация метода ближайшего соседа

Метод максимального правдоподобных расстояний

$$c^* = \arg \max_{c \in \{1, \dots, C\}} \log f \left(\rho(\mathbf{x}, \mathbf{x}_1), \dots, \rho(\mathbf{x}, \mathbf{x}_R) \mid W_c \right)$$

$$c^* = \arg \min_{c \in \{1, \dots, C\}} \sum_{r=1}^R \phi_c(r),$$

$$\phi_c(r) \approx \left(\rho(\mathbf{x}, \mathbf{x}_r) - \rho_{c, c_r} \right)^2 / \rho_{c, c_r}$$

- Savchenko A.V., Belova N.S. Unconstrained Face Identification Using Maximum Likelihood of Distances Between Deep Off-the-shelf Features, Expert Systems With Applications 108C (2018) pp. 170-182
- Savchenko A.V., Belova N.S., Savchenko L.V. Fuzzy Analysis and Deep Convolution Neural Networks in Still-to-video Recognition, Optical Memory and Neural Networks (Information Optics), 2018, vol. 27, no. 1, pp. 23-31
- Savchenko A.V., Belova N.S. Statistical testing of segment homogeneity in classification of piecewise-regular objects // International Journal of Applied Mathematics and Computer Science.– 2015.– Vol. 25, №4.– P. 915-925.
- Савченко А.В. Метод максимально правдоподобных рассогласований в задаче распознавания изображений на основе глубоких нейронных сетей // Компьютерная оптика.– 2017. – Т.41, №3.– С.422-430.



ТЕКУЩИЕ РЕЗУЛЬТАТЫ

Модификация вероятностной нейронной сети (ВНС)

Оценки Парзена в ВНС

$$c^* = \operatorname{argmax}_{c \in \{1, \dots, C\}} \frac{R(c)}{R} f(\mathbf{x}|W_c)$$

$$\hat{f}(\mathbf{x}|W_c) = \frac{1}{R(c)} \sum_{r=1}^{R(c)} K(\mathbf{x}, \mathbf{x}_r(c))$$

Проекционные оценки

$$\hat{f}_d(x_d|W_c) = \frac{1}{J+1} \sum_{j=0}^J \hat{f}_{j;d}(x_d|W_c)$$

$$\hat{f}_{J;d}(x_d|W_c) = \frac{1}{R(c)} \sum_{r=1}^{R(c)} D_J(x_d - x_{r;d}(c))$$

$$c^* = \operatorname{argmax}_{c \in \{1, \dots, C\}} \sum_{d=1}^D \log \sum_{j=-J}^J w_{j;d}(c) \cdot \psi_j(x_d)$$

$$w_{j;d}(c) = \frac{J+1-|j|}{(J+1)R(c)} \sum_{r=1}^{R(c)} \exp(ij\pi x_{j;d}(c))$$

- Савченко А.В. Тригонометрическая система функций в проекционных оценках плотности вероятности нейросетевых признаков изображений // Компьютерная оптика. – 2018. – Т. 42, No 1. – С. 149-158
- Savchenko A.V. Efficient Statistical Face Recognition Using Trigonometric Series and CNN Features, Proceedings of IEEE 23th International Conference on Pattern Recognition (ICPR 2018), pp. 3262-3267, 2018



ПРИКЛАДНЫЕ ЗАДАЧИ АНАЛИЗА ИЗОБРАЖЕНИЙ

ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ (1)

Распознавание пола и возраста на видео

Агрегация решения для каждого кадра на основе теории Демпстера-Шафера

$$DT_j^l = \frac{1}{N_j} \sum_{X \in \omega_j} P(l|X)$$

$$l^* = \operatorname{argmax}_{j=1, \overline{L}} \prod_{t=1}^T b_j(t).$$

Доверие (belief) Демпстера

$$b_j(t) = \frac{\Phi_j(t) \prod_{l \neq j} (1 - \Phi_l(t))}{1 - \Phi_j(t) \left[1 - \prod_{l \neq j} (1 - \Phi_l(t)) \right]}.$$

$$\Phi_j(t) = \frac{\left(1 + \left(\sum_{l=1}^L (DT_j^l - P(l|X(t))) \right)^{1/2} \right)^{-1}}{\sum_{i=1}^L \left(1 + \left(\sum_{l=1}^L (DT_i^l - P(l|X(t))) \right)^{1/2} \right)^{-1}}.$$

- Kharchevnikova A.S., Savchenko A.V. Neural Networks in Video-Based Age and Gender Recognition on Mobile Platforms, Optical Memory and Neural Networks (Information Optics), 2018, vol. 27, no. 4, pp. 246-259
- Kharchevnikova A.S., Savchenko A. V. Video-based age and gender recognition in mobile applications, Proceedings of the International Conference on Information Technology and Nanotechnology (ITNT). Session Image Processing and Earth Remote Sensing, 2018, CEUR-WS, vol. 2210, pp. 227-235.
- Kharchevnikova A.S., Savchenko A.V. (2018) The Video-Based Age and Gender Recognition with Convolution Neural Networks. In: Kalyagin V., Pardalos P., Prokopyev O., Utkina I. (eds) Computational Aspects and Applications in Large-Scale Networks. NET 2016. Springer Proceedings in Mathematics & Statistics, vol 247, pp. 37-46. Springer, Cham



ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ (2)

Кластеризация видеоданных

Совместно с Соколовой А.

- Исследованы способы агрегации векторов признаков каждого кадра
 - Проанализированы результаты различных методов кластеризации (в т.ч. Rank-order, dominant sets) для малого количества треков большого числа людей
-
- Sokolova A.D., Kharchevnikova A.S., Savchenko A.V. (2018) Organizing Multimedia Data in Video Surveillance Systems Based on Face Verification with Convolutional Neural Networks. In: van der Aalst W. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2017. Lecture Notes in Computer Science, vol 10716, pp 223-230. Springer, Cham
 - Sokolova A.D., Savchenko A.V. (2018) Cluster Analysis of Facial Video Data in Video Surveillance Systems Using Deep Learning. In: Kalyagin V., Pardalos P., Prokopyev O., Utkina I. (eds) Computational Aspects and Applications in Large-Scale Networks. NET 2016. Springer Proceedings in Mathematics & Statistics, vol 247, pp. 113-120. Springer, Cham
 - Sokolova A.D., Savchenko A. V. Data organization in video surveillance systems using deep learning, Proceedings of the International Conference on Information Technology and Nanotechnology (ITNT). Session Image Processing and Earth Remote Sensing, 2018, CEUR-WS, vol. 2210, pp. 243-250



ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ (3)

Распознавание эмоций групп людей

Совместно с Рассадиным А.

- Ансамбль нейронных сетей для участия в конкурсе EmotiW
 - Повышение эффективности за счет применения простых сетей, предобученных на распознавание эмоций для лиц малого размера (64x64)
-
- Rassadin A., Gruzdev A., Savchenko A. Group-Level Emotion Recognition using Transfer Learning from Face Identification. In Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17). ACM, 2017. P. 544-548
 - Tarasov A.V., Savchenko A.V. (2018) Emotion Recognition of a Group of People in Video Analytics Using Deep Off-the-Shelf Image Embeddings. In: van der Aalst W. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2018. Lecture Notes in Computer Science, vol 11179, pp 191-198. Springer, Cham



ПЕРСПЕКТИВЫ

Совместно с Рассадиным А.

- Распознавание лиц на видео с частичным привлечением учителя
- Распознавание сцен на фотографиях на основе комитета классификаторов и детекторов объектов

Совместно с Гречиным И.

- Анализ предпочтений пользователя по фото и видео на мобильном устройстве на основе нейросетевых детекторов объектов на изображениях

Совместно с Дёмочкиным К.

- Нейросетевая агрегация векторов признаков в наборе изображений при построении рекомендательных систем



СЖАТИЕ НЕЙРОННЫХ СЕТЕЙ



ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

Совместно с Рассадиным А.

- Сжатие сверточных сетей при распознавании эмоций

Совместно с Грачевым А. и Игнатовым Д. (НИУ ВШЭ, Москва)

- Сжатие рекуррентных нейронных сетей для языкового моделирования: матричные разложения (в том числе tensor-train) рекуррентных слоев и выходного полносвязного слоя
- Rassadin A. G., Savchenko A. V. Compressing deep convolutional neural networks in visual emotion recognition, Proceedings of the International Conference on Information Technology and Nanotechnology (ITNT). Session Image Processing, Geoinformation Technology and Information Security Image Processing (IPGTIS), 2017, CEUR-WS, vol. 1901, pp. 207-213.
- Rassadin A. G., Savchenko A. V. Deep neural networks performance optimization in image recognition, Proceedings of the 3 International Conference on Information Technologies and Nanotechnologies (ITNT), 2017, pp. 649-654.
- Grachev A.M., Ignatov D.I., Savchenko A.V. (2017) Neural Networks Compression for Language Modeling. In: Shankar B., Ghosh K., Mandal D., Ray S., Zhang D., Pal S. (eds) Pattern Recognition and Machine Intelligence. PReMI 2017. Lecture Notes in Computer Science, vol 10597, pp 351-357. Springer, Cham



ПЕРСПЕКТИВЫ

Рассадин А., Попова А.

- Структурное сжатие нейронных сетей для повышения вычислительной эффективности алгоритма распознавания



ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ

ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

Распознавание команд в системах голосового управления

Агрегация решения для каждого фрейма с представлением эталонных фонем в виде нечеткого множества

$$\mu(\mathbf{x}_r^*|\mathbf{x}(t)) \stackrel{def}{=} \hat{P}(\mathbf{x}_r^*|\mathbf{x}(t)) = \frac{\exp(-\lambda\rho_{KL}(\mathbf{x}(t), \mathbf{x}_r^*))p_r}{\sum_{k=1}^R \exp(-\lambda\rho_{KL}(\mathbf{x}(t), \mathbf{x}_k^*))p_k}$$

$$\mu_i(\mathbf{x}_r^*) = \frac{\exp(-\lambda\rho_{KL}(\mathbf{x}_i^*, \mathbf{x}_r^*))p_r}{\sum_{k=1}^R \exp(-\lambda\rho_{KL}(\mathbf{x}_i^*, \mathbf{x}_k^*))p_k}$$

$$\mu(r, t) = \mu_V(t)(\mathbf{x}_r^*)\mu(\mathbf{x}_r^*|\mathbf{x}(t)).$$

$$\mu_i(\mathbf{x}_r^*) = 1 - \prod_{j=1}^{J_i} (1 - P(\mathbf{x}_r^*|\mathbf{x}_{i,j}))$$

- Savchenko A.V. Sequential Three-Way Decisions in Efficient Classification of Piecewise Stationary Speech Signals. In: Polkowski L. et al. (eds) Rough Sets. IJCRS 2017. Lecture Notes in Computer Science/LNAI, vol 10314, pp. 264-277. Springer, Cham (2017)
- Savchenko V.V., Savchenko A.V. Information-theoretic analysis of efficiency of the phonetic encoding–decoding method in automatic speech recognition // Journal of Communications Technology and Electronics, 2016, Vol. 61, No. 4, pp. 430-435
- Savchenko A.V., Savchenko L.V. Towards the creation of reliable voice control system based on a fuzzy approach // Pattern Recognition Letters.– 2015.– Vol. 65.– P. 145-151.
- Savchenko A.V., Khokhlova Ya.I. About Neural-Network Algorithms Application in Viseme Classification Problem With Face Video in Audiovisual Speech Recognition Systems, Optical Memory and Neural Networks (Information Optics), 2014, vol. 23, no. 1, pp. 34–42



ПЕРСПЕКТИВЫ

- Модификация алгоритмов распознавания на основе «нормализованных» (gain-normalized) мер близости

$$d'(f, \hat{f}) \triangleq \min_{\lambda \geq 0} d(f, \lambda \hat{f}).$$

Совместно с Соколовым А.

- Распознавание речи с предметно-ориентированным словарем на основе глубоких нейронных сетей
- Распознавание атрибутов (пол, возраст) диктора по видео и выбор соответствующей акустической модели для дальнейшего распознавания речи

- Sokolov A., Savchenko A.V. Voice command recognition in intelligent systems using deep neural networks, Proceedings of IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI 2019), pp. 113-116, 2019



ЗАКЛЮЧЕНИЕ



ИССЛЕДОВАТЕЛЬСКИЕ ПРОЕКТЫ

И научные контакты

1. Visual Preference Prediction on Visual Data

- Николенко С.И., руководитель лаборатории искусственного интеллекта ПОМИ РАН-Samsung
- Мясников Е.В., доцент кафедры геоинформатики и информационной безопасности Самарского национального исследовательского университета
- Кузнецов А.В., доцент кафедры геоинформатики и информационной безопасности Самарского национального исследовательского университета
- Макаров И.А., ст. преп. НИУ ВШЭ (Москва)

2. НУГ «Анализ мультимедийных данных пользователей мобильных устройств»

- Малафеев А.Ю., доцент НИУ ВШЭ-НН
- Наши аспиранты и студенты: Соколов А., Гречихин И., Соколова А., Харчевникова А., Дёмочкин К., ...

3. Грант Президента 2017-2018

- Игнатов Д.И., доцент департамента анализа данных и искусственного интеллекта НИУ ВШЭ (Москва)
- Грачев А.М., исследователь в Samsung R&D



ПЕРСПЕКТИВЫ И ПРОБЛЕМЫ

Стремительное внедрение технологий AI

+ Потребность промышленных компаний в кадрах:

- Новые индустриальные партнеры
- новые проекты

- Потенциальные исполнители (сотрудники/студенты) проводят активные исследования не более 1-2 лет



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ