

Proposition d'un Postdoc

Titre : Extraction d'information dans des images scannées en vue de l'analyse de la consommation sur le marché hors domicile

Période : 1 mai 2020 au 30 octobre 2021 (18 mois)

Contact : A. Belaïd, LORIA, Equipe READ, France, abdel.belaid@loria.fr

Objectif

Il s'agit de créer un ensemble d'outils numériques, à destination des marques de l'agroalimentaire et des professionnels de la restauration hors domicile, permettant la promotion vers des consommateurs finaux et l'analyse de la consommation sur le marché hors domicile. A partir d'une simple photographie de son ticket de caisse de consommation hors domicile (CHD), l'utilisateur acquiert des points de fidélité à échanger contre des offres partenaires. Le système informatique à mettre en place comprend plusieurs étapes : acquisition de l'image du ticket à partir de son smartphone, localisation des lignes du texte, et extraction de l'information.

La littérature propose depuis quelques années des solutions pour certains secteurs d'activités, comme les débits de boissons, mais ces solutions sont trop restrictives à des types d'informations à extraire, ou encore dans le traitement des langues. On peut mentionner Mobishop [1] qui est une application de détection participative qui facilite le partage des prix des produits entre les participants, OCRdroid [2] qui s'appuie sur le moteur de reconnaissance optique de caractères Tesseract [3] et cherche à améliorer les problèmes de désalignement et de lumière insuffisante, Receiptlog [4] pour identifier les préférences d'achat spécialisées des utilisateurs afin de prévoir leur comportement futur, ou enfin l'application de Dikici et Saraclar [5] qui utilise des repères linguistiques pour corriger les textes glissants qui sont présentés dans les bulletins de nouvelles.

Par ailleurs, cette année, en marge de la conférence internationale ICDAR (International Conference on Document Analysis and Recognition), un groupe d'évaluation des tickets de caisse appelé SROIE (Scanned receipts OCR and key information extraction) a vu le jour [6]. Le rapport réalisé à l'issue de cette compétition indique que « bien que les tâches de localisation et de reconnaissance du texte semblent relativement faciles à aborder, il est intéressant d'observer la variété d'idées et d'approches proposées pour l'extraction d'informations. Selon les résultats des communications, nous pensons qu'il reste encore une marge pour améliorer les performances d'extraction d'informations ». Nous avons fait le même constat et proposons à la société VAZEE une solution plus performante de ses tickets de caisse.

Le sujet

Inspirés par les succès récents enregistrés par les modèles de deep learning et convaincus de l'apport de la linguistique dans ce cas, nous proposons une solution neuronale utilisant des modèles de langage. Pour la détection des lignes de texte, nous voudrions adapter la méthode EAST (Efficient and Accurate Scene Text Detector) [7] et le réseau connexionniste de proposition de texte (Connectionist Text Proposal Network (CTPN)) [8] qui ont été proposés pour détecter le texte dans des images naturelles.

EAST utilise un seul réseau de neurones pour prédire un mot ou un texte au niveau de la ligne. Il peut détecter le texte dans une orientation arbitraire avec des formes quadrilatérales. En 2017, cet algorithme a surpassé les méthodes de pointe. Il consiste en un réseau entièrement convolutionnel avec un état de fusion (merge) avec suppression non maximale.

CTPN explore les informations de contexte riches d'une image d'entrée, ce qui en fait un outil puissant pour détecter le texte dans différents formats de tickets. La structure CTPN est fondamentalement similaire à Faster R-CNN, mais avec l'ajout de la couche LSTM. Le modèle de réseau comprend principalement trois parties: l'extraction de caractéristiques par VGG16, le LSTM bidirectionnel et la régression du cadre de sélection.

Aussi bien la langue que les tournures linguistiques seront intégrés dans ces réseaux. Ainsi, on évitera les systèmes classiques qui nécessitent de trouver d'abord la langue, puis d'appliquer en conséquence les règles concomitantes pour trouver l'information. Un grand nombre d'échantillons de documents annotés serait nécessaire pour réaliser cette étape.

L'information à extraire correspond à un mot ou un groupe de mots réunis dans un assemblage qu'on appelle « contexte ». Le réseau choisi devra apprendre ces assemblages et en prédire lors de la

phase de reconnaissance. Pour leur représentation, on pourra soit utiliser une représentation vectorielle comme Word2Vec soit un graphe de voisinage, type GAN ou GraphGAN [9].

La bibliographie

- [1] Sehgal, S., Kanhere, S. S. and Chou, C. T., "Mobishop: Using mobile phones for sharing consumer pricing information," Proc. Conference on Distributed Computing in Sensor Systems, Santorini, Greece (2008).
- [2] Joshi, A., Zhang M., Kadmwala, R., Dantu, K., Poduri, S. and Sukhatme, G., "OCRdroid : A Framework to Digitize Text Using Mobile Phone," Proc. ICST International Conference on Mobile Computing, Applications, and Services, (2009).
- [3] Smith, R., "An overview of the Tesseract OCR engine," Proc. International Conference on Document Analysis and Recognition, Brazil, (2007).
- [4] Tokunaga, S., Matsumoto, S. and Nakamura, M., "Receiptlog: A consumer-oriented lifelog service for storing and reviewing daily receipts", IEICE Technical Report, vol. 111, no.107, 23–28 (2011).
- [5] Dikici, E., and Saraçlar, M., "Sliding text recognition in broadcast news," Proc. IEEE 16th SignalProcessing and Communications Applications Conference (SIU), (2008).
- [6] Z. Huang, K. Chen, J. He, X. Bai, D. Karatzas, S. Lu and C. V. Jawahar, ICDAR2019 Competition on Scanned Receipt OCR and Information Extraction, ICDAR 2019, pp. 1516-1520.
- [7] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, Jiajun Liang, EAST: An Efficient and Accurate Scene Text Detector, CVPR 2017, pp. 5551-5560.
- [8] Tian, Zhi, Weilin Huang, Tong He, Pan He, and Yu Qiao, Detecting Text in Natural Image with Connectionist Text Proposal Network, ECCV (2016).
- [9] Hongwei Wang, Jia Wang, Jialin Wang, Miao Zhao, Weinan Zhang, Fuzheng Zhang, Xing Xie, Minyi Guo, GraphGAN: Graph Representation Learning with Generative Adversarial Nets, The 32nd AAAI Conference on Artificial Intelligence (AAAI 2018), arXiv:1711.08267.