

One-shot learning van gebaren in een convolutioneel neuraal netwerk

Jasper Vaneessen

Promotor: prof. dr. ir. Joni Dambre

Begeleider: Lionel Pigou

Masterproef ingediend tot het behalen van de academische graad van
Master of Science in de industriële wetenschappen: informatica

Vakgroep Elektronica en Informatiesystemen
Voorzitter: prof. dr. ir. Rik Van de Walle
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2016-2017



Toelating tot bruikleen

“De auteur geeft de toelating deze scriptie voor consultatie beschikbaar te stellen en delen van de scriptie te kopiëren voor persoonlijk gebruik.

Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting de bron uitdrukkelijk te vermelden bij het aanhalen van resultaten uit deze scriptie.”

Jasper Vaneessen, juni 2017

Voorwoord

Jasper Vaneessen, juni 2017

Overzicht

One-shot learning van gebaren in een convolutioneel neuraal netwerk

Jasper Vaneessen

Promotor: Prof. Dr. Ir. Joni DAMBRE
Begeleider: Ir. Lionel PIGOU

Masterproef ingediend tot het behalen van de academische graad van
MASTER OF SCIENCE IN DE INDUSTRIËLE WETENSCHAPPEN:
INFORMATICA

Vakgroep Elektronica en Informatiesystemen
Voorzitter: Prof. Dr. Ir. Rik VAN DE WALLE

Faculteit Ingenieurswetenschappen & Architectuur
Academiejaar 2016-2017 Universiteit Gent

Trefwoorden

gebarenherkenning, convolutionele neurale netwerken, machine learning, deep learning,
one-shot learning

Inhoudsopgave

Toelating tot bruikleen	ii
Voorwoord	iii
Overzicht	iv
Overzicht	iv
Extended abstract	v
Inhoudsopgave	v
Afkortingen	vi
1 Inleiding	1
1.1 Gebarentaal	1
1.2 Automatische gebarentaalherkenning	2
1.2.1 Gebarenherkenning	2
1.2.2 Gebarensegmentatie	2
1.3 One-shot learning	2
1.4 Doelstelling	2
2 Technische aspecten	3
2.1 Machine Learning	3
2.2 Artificieel neurale netwerk	4
2.3 Convolutioneel neurale netwerk	4
2.4 Hyperparameters	4
Bibliografie	5

Afkorting

ANN Artificiëel Neuraal Netwerk

CNN Convolutioneel Neuraal Netwerk

Hoofdstuk 1

Inleiding

1.1 Gebarentaal

Gebarentaal is in de eerste plaats een taal. Taal is een begrip dat moeilijk te definiëren valt en de meeste pogingen hiertoe beperken zich tot gesproken taal. Een definitie die ook voor gebarentaal kan gebruikt worden is: een natuurlijk ontstaan communicatiemiddel waarmee je kan communiceren over alles wat je denkt, ziet, voelt en droomt. [?]

Gesproken taal en gebarentaal verschillen in de manier waarop gecommuniceerd wordt: oraal-auditief tegenover gestueel-visueel. Door middel van hand-, hoofd- en armbewegingen wordt een woord uitgesproken en vervolgens visueel waargenomen.

Een gebarentaal ontstaat, net zoals een gesproken taal, spontaan en natuurlijk door contact tussen mensen. Net door deze spontane ontwikkeling is er geen universele gebarentaal. Evenals we verschillende gesproken talen en dialecten kennen per land of regio zijn er ook verschillende gebarentalen [?]. In Nederland is er bijvoorbeeld de Nederlandse Gebarentaal (NGT) en in België de Vlaamse Gebarentaal (VGT) en de Waalse Gebarentaal (la Langue des Signes de Belgique Francophone, LSFB).

Een gebarentaal heeft een eigen grammatica en lexicon. Het lexicon of de gebarenschat is de verzameling van alle woorden of gebaren in de taal. Het lokale gebarenschat moet

volledig onafhankelijk vna het lokale woordenschat worden beschouwd.

Bepaalde woorden uit de ene taal kunnen niet eenduidig vertaald worden in een andere taal. Het woord "gezelligheid" kent bijvoorbeeld geen Engelse vertaling en voor het Duitse "fingerspitzengefühl" hebben we in de Nederlandse taal ook geen alternatief.

Tussen een gebarentaal en een gesproken taal geldt dezelfde verhouding.

Visuele taal hand, hoofd, lichaamsbewegingen eigen woordenschat en grammatica een volwaardige taal geen een op een vertaling van woord naar gebaar

regionale verschillen, vlaamse gebarentaal, nederlandse gebarentaal zijn er uit zichzelf gekomen, pogingen tot standaardisatie vele varianten moeilijke communicatie onderling communicatie met horenden via schrift,

1.2 Automatische gebarentaalherkenning

1.2.1 Gebarenherkenning

1.2.2 Gebarensegmentatie

1.3 One-shot learning

1.4 Doelstelling

Hoofdstuk 2

Technische aspecten

2.1 Machine Learning

Leren is een veelzijdig fenomeen dat bestaat uit verschillende processen: het verkrijgen van declaratieve kennis, het ontwikkelen van motorische en cognitieve vaardigheden door instructie en ervaring, het organiseren van nieuwe kennis in algemene representaties en het ontdekken van nieuwe feiten via observatie en experimentatie.

Sinds het begin van het computertijdperk proberen onderzoekers het menselijk leren na te bootsen en deze processen te vertalen naar de informatietheorie. Het machinaal leren is nog steeds een erg uitdagend doel in de kunstmatige intelligentie (KI).

We kunnen zeggen dat een computerprogramma of machine leert als het zijn performantie op een bepaalde taak verbetert met ervaring [Carbonell et al., 1983]

In het geval van dit onderzoek is deze taak het classificeren van gebaren. Het model zal dus leren

2.2 Artificiëel neurale netwerk

2.3 Convolutioneel neurale netwerk

2.4 Hyperparameters

Bibliografie

- [Ba et al., 2015] Ba, J., Swersky, K., Fidler, S., and Salakhutdinov, R. (2015). Predicting Deep Zero-Shot Convolutional Neural Networks using Textual Descriptions. *arXiv:1506.00511 [cs]*. arXiv: 1506.00511.
- [Braun et al., 2009] Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C. (2009). Motor Task Variation Induces Structural Learning. *Current Biology*, 19(4):352–357.
- [Caelles et al., 2016] Caelles, S., Maninis, K.-K., Pont-Tuset, J., Leal-TaixÃ©, L., Cremers, D., and Van Gool, L. (2016). One-Shot Video Object Segmentation. *arXiv:1611.05198 [cs]*. arXiv: 1611.05198.
- [Carbonell et al., 1983] Carbonell, J. G., Michalski, R. S., and Mitchell, T. M. (1983). An Overview of Machine Learning. In Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., editors, *Machine Learning*, Symbolic Computation, pages 3–23. Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-12405-5_1.
- [Ciresan et al., 2012] Ciresan, D. C., Meier, U., and Schmidhuber, J. (2012). Transfer learning for Latin and Chinese characters with deep neural networks. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE.
- [Cooper et al., 2012] Cooper, H., Ong, E.-J., Pugeault, N., and Bowden, R. (2012). Sign language recognition using sub-units. *Journal of Machine Learning Research*, 13(Jul):2205–2231.

- [Hoffman et al., 2013] Hoffman, J., Tzeng, E., Donahue, J., Jia, Y., Saenko, K., and Darrell, T. (2013). One-Shot Adaptation of Supervised Deep Convolutional Models. *arXiv:1312.6204 [cs]*. arXiv: 1312.6204.
- [Ji et al., 2013] Ji, S., Xu, W., Yang, M., and Yu, K. (2013). 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):221–231.
- [Kadrev et al., 2016] Kadrev, G., Kostadinov, G., and Ruskov, P. (2016). Expansion of a CNN-based image classifier’s scope using transfer learning and k-NN. Technical report. In *2016 IEEE 8th International Conference on Intelligent Systems (IS)*, pages 764–770.
- [Karpathy et al., 2014] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732.
- [Lake et al., 2015] Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338.
- [Lima et al., 2017] Lima, E., Sun, X., Dong, J., Wang, H., Yang, Y., and Liu, L. (2017). Learning and Transferring Convolutional Neural Network Knowledge to Ocean Front Recognition. *IEEE Geoscience and Remote Sensing Letters*, 14(3):354–358.
- [Michalski et al., 2013] Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (2013). *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media.
- [Oquab et al., 2014] Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1717–1724.

- [Pan and Yang, 2010] Pan, S. J. and Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.
- [Pigou,] Pigou, L. Gebarentaalherkenning met convolutionele neurale.
- [Samuel, 1959] Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3):210–229.
- [Santoro et al., 2016] Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. (2016). One-shot Learning with Memory-Augmented Neural Networks. *arXiv preprint arXiv:1605.06065*.
- [Wu et al., 2016] Wu, D., Pigou, L., Kindermans, P.-J., Nam, L. E., Shao, L., Dambre, J., and Odobez, J.-M. (2016). Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition.
- [Wu and Shao, 2014] Wu, D. and Shao, L. (2014). Deep dynamic neural networks for gesture segmentation and recognition. In *Workshop at the European Conference on Computer Vision*, pages 552–571. Springer.
- [Wu et al., 2012] Wu, D., Zhu, F., and Shao, L. (2012). One shot learning gesture recognition from RGBD images. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 7–12.
- [Zaki and Shaheen, 2011] Zaki, M. M. and Shaheen, S. I. (2011). Sign language recognition using a combination of new vision based features. *Pattern Recognition Letters*, 32(4):572–577.
- [Zeiler and Fergus, 2014] Zeiler, M. D. and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision - ECCV 2014*, Lecture Notes in Computer Science, pages 818–833. Springer International Publishing. DOI: 10.1007/978-3-319-10590-1_53.