MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220079

Using Deep Learning to Improve Phenotyping from Clinical Reports

Marc Vincent[a], Maxime Douillet[a], Ivan Lerner[b], Antoine Neuraz[b], Anita Burgun[b,c,d], Nicolas Garcelon[a]

[a] *Institut Imagine, Paris Descartes University-Sorbonne Paris Cité, Paris, France*
[b] *INSERM UMR1138, Centre de Recherche des Cordeliers, Team 22, Paris, France*
[c] *Department of Medical Informatics, Necker-Enfants Malades Hospital, Assistance Publique - Hôpitaux de Paris (AP-HP), Paris, France*
[d] *Paris Descartes University Sorbonne Paris Cité, Paris, France*

Abstract

With the development of clinical databases and the ubiquity of EHRs, physicians and researchers alike have access to an unprecedented amount of data. Complexity of the available data has also increased since clinical reports are also included and require frameworks with natural language processing capabilities in order to process them and extract information not found in other types of documents. In the following work we implement a data processing pipeline performing phenotyping, disambiguation, negation and subject prediction on such reports. We compare it to an existing solution routinely used in a children's hospital with special focus on genetic diseases. We show that by replacing components based on rules and pattern matching with components leveraging deep learning models and fine-tuned word embeddings we obtain performance improvements of 7%, 10% and 27% in terms of F1 measure for each task. The solution we devised will help build more reliable decision support systems.

Keywords:

Natural Language Processing, Phenotype, Data Warehousing, Deep Learning

Introduction

The extraction of information from medical reports becomes essential to allow a computational representation of patients. Particularly in the context of genetic diseases and patients in diagnostic errancy, automated phenotype extraction allows to produce a fine-grained representation of patients, to facilitate the stratification of those patients and to normalize patients' EHR to apply clinical decision support system.

In order to improve the Necker Hospital data warehouse framework (DrWarehouse [1]), we sought to produce a deep learning-based method that would allow us to properly extract phenotyping information from clinical reports. More precisely, we aimed at denoising and qualifying phenotypes extracted by a pattern-based method already used in DrWarehouse. We focused on three aspects: shallow disambiguation, negation and subject prediction. The disambiguation problem stems from the fact that an expression may have different meanings in different contexts, for example the French word 'grippe' might refer to a request a vaccine against the flu (i.e. a treatment) or to the flu itself (i.e. a phenotype). Conflating the two meanings would result in a misrepresentation of the concepts related to the term 'grippe' in the analysed clinical reports and may therefore impact the downstream tasks.

The negation problem on the other hand stems from the fact that a (true) phenotype may be referred to in an clinical report to assert its presence in a patient, or on the contrary to negate it. Once again, conflating the two possibilities might result in very different interpretations in the downstream analysis. Finally, the subject prediction problem amounts to discriminating mentions of observations that describe the patient (to whom the report belongs) vs. mentions that describe members of its family. The latter two problems can be referred to in the literature as meta-annotation or attribute prediction [2], the annotation of negation has been included in the creation of corpora such as Bioscope [3] or Merlot [4].

High throughput phenotyping may cover and describe quite different tasks. Part of the published solutions for phenotyping focus on specific phenotypes of interest that can be predicted using structured data and billing codes or clinical reports and occurrences of phenotype related terms combined to classification rules or neural networks [5]. Predicting said phenotype of interest can later be used to constitute cohorts for further studies of a given disease. This approach has been partly automated in Zhang et al. [6] by creating a pipeline adaptable to any phenotype by leveraging chart-review-defined gold standard. Yang et al. [5] focuses on studying the influence of the combination of word, sentence embedding and CNNs in specific phenotype classification performances.

Another set of solutions focus on extracting all the phenotypes terms occurring in the clinical reports. This can be seen as both more specific (it can be a step of aforementioned pipelines) and more general (we do not focus on one phenotype of interest). These descriptors can later be used for querying a database for patients fitting a certain profile under investigation (when hypotheses and search for evidence underpinning them are still being made) and without constraining the search on previous assumptions or known cases. They can also be used as a prerequisite to perform a phenome-wide association study as noted in Robinson et al. [7]. This is the approach adopted in DrWarehouse [1] or DeepPhe [8].

More generally, the phenotype mention problem also can be seen as a special case of either Concept Extraction or Named Entity Recognition tasks applied to phenotypes. The former task has seen the development of a number of solutions leveraging existing medical thesauri to search for mentions of the terms it contains in clinical notes or other documents, with the optional use of shallow syntactic parsing and fuzzy matching, e.g. cTakes [9] or QuickUMLS [10]. For the latter task a class of methods based on corpus annotation and supervised learning have been successfully used in the biomedical context,

such methods can leverage deep learning models such as BiSTM-CRF [11]. These methods can benefit from use of silver standard annotations and transfer learning to alleviate the annotation burden as noted by Giorgi and Bader [12], who use already existing annotation pipelines to produce such silver standard annotations or more recently in PhenoTagger [13].

Our method adopts an effective approach based on leveraging robust pattern-based search of phenotypic terms included in an existing medical thesaurus (UMLS) and flexible machine learning based attribute prediction on these terms. It is also a method that has been applied to clinical reports redacted in French, which share specific constraints with similar works on other languages than English such as scarcer resources to leverage, and the difficulty to adapt existing tools such as cTakes as noted in Névéol et al. [14].

Methods

We relied on two datasets of 2222 and 900 clinical reports annotated in two steps, first using an automated exact match and pattern-based method to detect potential phenotypes, then manually revising the annotations and augment them with predefined attributes (i.e., negation, subject). To learn from this dataset we relied on deep learning models atop fine-tuned word embeddings.

Corpora

The corpora were made by sampling free-text Electronic Health Records from the Necker's Children hospital's database (containing $2{,}5.10^6$ documents in total). The sampling of the documents was stratified on the provider unit to ensure the representativity of the corpus. A maximum of one document per patient could be selected in the sample. The annotation process was carried out by two experts who were tasked with checking and enriching phenotype annotations made in an automated fashion by an exact match and pattern-based method (as implemented in DrWarehouse [1]).

First, the preliminary automated annotation consisted in using the French portion of the UMLS metathesaurus to get terms belonging to 10 semantic groups (provided as part of the UMLS hierarchy) deemed to relate to phenotypes [1]. Then, using a brat annotation server [15] the experts had to report both true and false positives (i.e. expressions matching a UMLS phenotype term, that either did or did not correspond to a phenotype in context) as well as indicating if the presence of the phenotype was asserted or negated.

Given the scarcity of the family annotations it was decided to constitute a separate enriched dataset for this specific task: the collection of clinical reports from Necker's children hospital was searched for documents containing patterns suggesting family related mentions using the method previously implemented in DrWarehouse.

In every case, the quality of the annotations was evaluated by having two annotators performing redundant annotations on a subset of documents (n = 100 and n=60, respectively).

Annotation guidelines for phenotype disambiguation were made of five major rules, accompanied by a set of specific examples and terms, the five major rules where:

1. that the mention of detected term was a description of the physical and / or mental state of the patient
2. that the mention was a qualitative statement (i.e., excluding measures requiring threshold interpretation).
3. That the extracted terms had the same meaning than its mention in context
4. That the mention could describe a healthy state or a pathological one
5. That the extracted text was self-sufficient in meaning

Annotation rules regarding negation and subject were more succinct: Were flagged as negated mentions that appear in the text to denote the absence of the characteristic or lack of observation it describes. Were flagged as patient the mentions relating to the patient to which the clinical report belongs, and as family the one relating to its family. As an additional rule, it was decided that reports referring to a foetus were attributed to it as a patient.

Table 1 - Upper half of the table: Count of occurrences of the UMLS terms (and tokens) for each classification problem. In parentheses the positive count corresponds to the positive class (i.e., true phenotype, negated, patient mentions). When two lines are available the first one corresponds to counts in the test sets and the second one to full corpus counts (cf. main text). Cohen's Kappa is also reported.

Task	Terms (positives)	Term tokens (positives)	Kappa
Disamb.	6664 (5499) 12231 (10176)	10591 (8996) 19447 (16654)	0.81
Negation	7092 (1214) 13517 (2231)	11169 (1799) 20733 (3279)	0.92
Subject	5630 (4832)	8388 (6954)	0.95

Pre-Processing & UMLS terms search

Clinical reports were tokenized using Spacy (v2.3.2) French model (fr_core_news_sm). Counts of tokens in Table 1 refer to this tokenization process. UMLS terms were submitted to the same tokenization procedure and a full-text search for UMLS terms mentions was performed by aligning tokens of the UMLS terms to tokens of the clinical report. All documents and terms were submitted to lower-casing and accents removal based on a predefined dictionary.

Embeddings

A first set of embeddings was obtained using a skip gram fastText model [16] set to produce vectors of size 300 and trained with default parameters on a collection of 2.5 million clinical reports from the Necker-Enfants Malades hospital.

A second set of embeddings was obtained by fine-tuning a CamemBERT model [17] on the same set of clinical reports. CamemBERT is itself a RoBERTA model trained on a large French corpus of documents scrapped from the web. The tested model produces vectors of size 768. Compared to fastText, CamemBERT produces contextual embeddings that display different vector representations for the same token depending on the context it appears in. The tokenization itself relies on a sentence piece tokenizer, aligning the sentence piece tokens to spacy tokens was achieved by using the Flair library [18]. During the fine-tuning procedure full size reports were truncated -if necessary- to the limit of 512 tokens. The procedure, which took one week using three NVIDIA RTX2080Ti graphics cards, was carried out using Huggin Face's Transformers library. At inference time we also relied on the Flair library to overcome the limit of 512 tokens imposed by the naive CamemBERT model.

Table 2 - Performances of pattern based DrWarehouse (DrWH) compared to the classification provided by the selected word-embeddings and deep learning models (DrWH+). The metrics are computed on 10 mutually exclusive test sets. Measures in boldface are the highest nominal values.

Task	Model	Precision	Recall	F1	NPV	Specificity	F1n
Disambiguation	DrWH	82.5 +/- 1.6	**100**	90.4 +/- 1.0	NR	NR	NR
	DrWH+FT	**97.1 +/- 1.2**	98.2 +/- 0.7	**97.6 +/- 0.8**	**90.7 +/- 4.1**	86.1 +/- 5.8	**88.3 +/- 4.3**
	DrWH+CA	**97.1 +/- 1.5**	97.9 +/- 1.0	97.5 +/- 0.9	89.5 +/- 5.1	**86.4 +/- 6.2**	87.8 +/- 4.3
Negation	DrWH	78.1 +/- 4.0	93.0 +/- 2.0	84.9 +/- 2.9	98.5 +/- 0.5	94.6 +/- 1.2	96.5 +/- 0.8
	DrWH+FT	92.7 +/- 2.1	95.0 +/- 2.7	93.8 +/- 1.7	98.9 +/- 0.6	**98.5 +/- 0.4**	98.7 +/- 0.4
	DrWH+CA	**93.0 +/- 2.8**	**96.3 +/- 1.8**	**94.6 +/- 1.8**	**99.2 +/- 0.4**	**98.5 +/- 0.6**	**98.8 +/- 0.4**
Subject	DrWH	41.9 +/- 8.0	**88.0 +/- 7.1**	56.5 +/- 8.2	97.6 +/- 1.2	79.7 +/- 2.5	87.8 +/- 1.8
	DrWH+FT	84.0 +/- 8.4	78.6 +/- 9.8	80.5 +/- 5.0	96.5 +/- 1.8	97.3 +/- 1.9	96.9 +/- 0.9
	DrWH+CA	**87.3 +/- 9.2**	81.1 +/- 12.0	**83.0 +/- 5.9**	**96.9 +/- 1.8**	**97.9 +/- 1.6**	**97.4 +/- 0.8**

Models

The models we tested are recurrent neural networks (RNN) of two types: either Gated Recurrent Units (GRU) [19] or Long Short Term Memory (LSTM) [20]. Both are successful in the context of Natural Language Processing although the former is computationally simpler than the latter. Both also implement roughly the same principle, explicitly and selectively controlling how hidden states are overwritten based on new input and past states. Both models were trained with carefully tuned parameters and using a weighting scheme based on token labels (awarding more weight to the rarer class, based on the ratio between class frequencies and a parameter controlling how much the ratio is taken into account).

Training & Evaluation

We assessed the improvements brought by our models by comparing them to a previous exact-match and pattern-based extraction method, with respect to three problems: disambiguating between true and false phenotype mentions (disambiguation), stating if the mention was used to assert the presence of the phenotype or to deny it (negation), and if the mention can be attributed to the patient concerned by clinical report or her/his family (subject).

The pattern-based methods are reimplementations in the python programming language of DrWarehouse clinical extraction procedures which are used routinely at Necker's Children hospital to process patients' clinical reports.

The three problems were cast as a binary prediction problem on each token, then predictions at the token level were grouped at the entity level with a majority vote rule (i.e., if most of the tokens of an entity were predicted to be part of the positive class then the entity was considered to belong to the positive class itself, to the negative class otherwise).

Training was done using various combinations of parameters, including the size of the hidden layer of the RNN cells, dropout rates, number of hidden layers in the RNN, learning rate, and negative vs positive class weighting scheme.

The evaluation of performances was done using 10 partitions of the dataset in train / validation / test sets. The performance measures are reported for the mutually exclusive test sets, after selection of the parameters (and early stopping) on the validation set (see Table 2). Reported metrics are precision, recall, F1 (the harmonic mean of precision and recall with equal weights), Negative Predictive Value (NPV), specificity, and a measure we dubbed F1n: the harmonic mean of NPV and specificity. For the phenotype and negation prediction problems the test sets were taken from an earlier partition of the corpora, when roughly half of the annotation was made, hence the counts of terms and tokens are smaller when considering the test sets w.r.t. the full dataset.

Results

The proposed deep-learning based models improve over the pattern method as can be seen in Table 2. For disambiguation, negation and subject tasks the F1-score is increased by 7, 9% and 24% respectively when comparing the DrWH pattern method to the one including recurrent neural networks atop FastText embeddings (DrWH+FT for short). Replacing fastText embeddings by CamemBERT contextual embeddings (DrWH+CA) brings again an improvement in performances for Negation and Subject classification tasks, but of a smaller magnitude (bringing the total nominal improvements to 10%, and 27% respectively).

Looking more in detail at the disambiguation problem we may note that 17.5% of the terms the base matching method detects are not actual phenotypes. We may also see that the gain in F1 is made of a larger gain (15%) in precision and a small loss (2%) in recall. It is important to recall that in the disambiguation task, the positive class (phenotype rightfully detected by the match method) is the dominant class (ratio 5 to 1), therefore the small loss in recall should be contrasted with the fact that it allows us to correctly detect 86% of the tokens wrongly assigned to the phenotype class by the pattern method (specificity).

The negation prediction problem benefits in all respects from the additional deep learning step, with all the performance metrics being better with that step, but the most important improvement is seen for the subject prediction problem with an overall increase of 27 points in F1.

It is however also for that task that the deep learning models seem to be struggling the most (reaching a lower nominal F1 value and displaying high variability across folds), indicating that the problem may be the harder of the three and/or that the size of the corpus may be a limiting factor. It is also interesting to note that the pattern method performs better at the recall level, but is much less successful at the precision level.

Indeed, the pattern-based methods rely on the co-occurrence in syntagms of an exhaustive set of words that broadly refer to family and the targeted phenotype mentions. So, while it captures a great deal of mentions it is also subject to a frailty induced by wrong sense attribution or wrong relation attribution. E.g., seen cases comprises the erroneous understanding of the French word 'fils' as meaning son when it actually refers to a surgical thread (same French spelling), or 'maternelle' as the French adjective referring to a property of the

mother when it is actually a reference to kindergarten (again with the same spelling).

Discussion

With this work we improved an existing NLP phenotyping pipeline for French clinical reports using deep learning models. The best model for each task is made available through dockerized microservices that can be queried to augment documents contained in the DrWarehouse database. In turn the additional steps will allow for more precise information retrieval for clinicians and researchers using the DrWarehouse database.

We also showed that on the tested prediction problems and corpora, using a transformer-based contextual embedding does improve measured performances (for two tasks, while leaving performances relatively untouched for the third one). This is consistent with several results in the NLP literature although it should be noted that it is not always the case as in [2].

For the subject prediction task, it was decided to assign clinical reports mentioning a foetus and pregnancy to the foetus. Although this was a decision consistent with the perceived usage of clinical reports in our context (i.e., following mainly the health of the foetus in a hospital dedicated to children's health), this poses clear limitations to the exploitation of those reports. In ulterior versions of the datasets and models we plan to add specific rules for that case, including a multi referential view of the clinical report (both mother and foetus/child as patients).

Another design choice was to annotate pre-detected terms in order to limit the annotation burden. This implies that the recall for phenotype detection is currently capped by the pattern method which is used prior to the attribute prediction.

Future iterations of the system may instead include detection using deep learning based NER. This would allow for robustness against spelling mistakes, as well as exotic variations in terms names not contained in the used lexicons. In turn such a system would also have a greater need to perform term normalization as in similar systems such as PhenoTagger [13]. Other planned enhancements include the addition of other attributes prediction (e.g., hypothesis, temporality) as well as gene and mutation detection.

As the devised models were tested on documents originating from one hospital, transferability across hospitals was not tested. In order to limit the source bias, we took care of annotating relatively large sets of documents, and sampling these sets from a various array of sources among the ones at our disposal. When sharing the models with other facilities, care will be taken to quantify the extent of the possible bias.

Finally, it should also be noted that both inference and fine-tuning times are greatly increased (roughly one order of magnitude) when using transformer embeddings. Care should therefore be taken to contrast gained performances with increased requirements when considering the used embeddings. In our case, two of the three models we selected for production are using a CamemBERT embedding.

Conclusions

The presented work enhances the natural language processing and information extraction capabilities of an information system used routinely at a children's hospital. It leverages in-house developed corpora, deep learning models and word embeddings to enhance phenotype extraction and contextualization. We evaluated the performance improvements realized to 7%, 10% and 27% of F1 measure respectively for the three prediction tasks presented here (disambiguation, negation, subject).

The developed system which augments patients' EHRs with more precise information will contribute to better characterization of patients and provide a deeper understanding of their commonalities. Further enhancements of the system will include augmenting contextualization by adding hypothesis and temporality extraction, as well as gene and mutation detection. We also consider introducing deep-learning based NER for the detection step coupled with further normalization.

Acknowledgements

This work was supported by State funding from The French National Research Agency (ANR) under "Investissements d'Avenir" programs (Reference: ANR-10-IAHU-01) and C'IL-LICO project (Reference: ANR-17-RHUS-0002).

References

[1] N. Garcelon, A. Neuraz, R. Salomon, H. Faour, V. Benoit, A. Delapalme, A. Munnich, A. Burgun, and B. Rance, A clinician friendly data warehouse oriented toward narrative reports: Dr Warehouse, J. Biomed. Inform. 80 (2018). doi:10.1016/j.jbi.2018.02.019.

[2] Z. Kraljevic, D. Bean, A. Mascio, L. Roguski, A. Folarin, A. Roberts, R. Bendayan, and R. Dobson, MedCAT -- Medical Concept Annotation Tool, ArXiv191210166 Cs Stat. (2019). http://arxiv.org/abs/1912.10166 (accessed May 7, 2021).

[3] V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik, The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes, BMC Bioinformatics. 9 Suppl 11 (2008) S9. doi:10.1186/1471-2105-9-S11-S9.

[4] L. Campillos, L. Deléger, C. Grouin, T. Hamon, A.-L. Ligozat, and A. Névéol, A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMSI annOtated Text corpus (MERLOT), Lang. Resour. Eval. 52 (2018) 571–601. doi:10.1007/s10579-017-9382-y.

[5] Z. Yang, M. Dehmer, O. Yli-Harja, and F. Emmert-Streib, Combining deep learning with token selection for patient phenotyping from electronic health records, Sci. Rep. 10 (2020). doi:10.1038/s41598-020-58178-1.

[6] Y. Zhang, T. Cai, S. Yu, K. Cho, C. Hong, J. Sun, J. Huang, Y.-L. Ho, A. Ananthakrishnan, Z. Xia, S. Shaw, V. Gainer, V. Castro, N. Link, J. Honerlaw, S. Huang, D. Gagnon, E. Karlson, R. Plenge, and K. Liao, High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP), Nat. Protoc. 14 (2019) 1–19. doi:10.1038/s41596-019-0227-6.

[7] J. Robinson, W.-Q. Wei, D. Roden, and J. Denny, Defining Phenotypes from Clinical Data to Drive Genomic Research, Annu. Rev. Biomed. Data Sci. 1 (2018) 1–24. doi:10.1146/annurev-biodatasci-080917-013335.

[8] G.K. Savova, E. Tseytlin, S. Finan, M. Castine, T. Miller, O. Medvedeva, D. Harris, H. Hochheiser, C. Lin, G. Chavan, and R.S. Jacobson, DeepPhe: A Natural Language Processing System for Extracting Cancer Phenotypes from Clinical Records, Cancer Res. 77 (2017) e115–e118. doi:10.1158/0008-5472.CAN-17-0615.

[9] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, J. Am. Med. Inform. Assoc. JAMIA. 17 (2010) 507–513. doi:10.1136/jamia.2009.001560.

[10] L. Soldaini, QuickUMLS: a Fast, Unsupervised Approach for Medical Concept Extraction, in: 2016.

[11] M. Habibi, L. Weber, M. Neves, D. Wiegandt, and U. Leser, Deep learning with word embeddings improves biomedical named entity recognition, Bioinforma. Oxf. Engl. 33 (2017) i37–i48. doi:10.1093/bioinformatics/btx228.

[12] J. Giorgi, and G. Bader, Transfer learning for biomedical named entity recognition with neural networks, Bioinforma. Oxf. Engl. 34 (2018). doi:10.1093/bioinformatics/bty449.

[13] L. Luo, S. Yan, P.-T. Lai, D. Veltri, A. Oler, S. Xirasagar, R. Ghosh, M. Similuk, P.N. Robinson, and Z. Lu, PhenoTagger: a hybrid method for phenotype concept recognition using human phenotype ontology, Bioinformatics. (2021) btab019. doi:10.1093/bioinformatics/btab019.

[14] A. Névéol, H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum, Clinical Natural Language Processing in languages other than English: Opportunities and challenges, J. Biomed. Semant. 9 (2018) 12. doi:10.1186/s13326-018-0179-8.

[15] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii, BRAT: A Web-Based Tool for NLP-Assisted Text Annotation, in: Proc. Demonstr. 13th Conf. Eur. Chapter Assoc. Comput. Linguist., Association for Computational Linguistics, USA, 2012: pp. 102–107.

[16] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, Bag of Tricks for Efficient Text Classification, CoRR. abs/1607.01759 (2016). http://arxiv.org/abs/1607.01759.

[17] L. Martin, B. Muller, P.J. Ortiz Suárez, Y. Dupont, L. Romary, É. de la Clergerie, D. Seddah, and B. Sagot, CamemBERT: a Tasty French Language Model, in: Proc. 58th Annu. Meet. Assoc. Comput. Linguist., Association for Computational Linguistics, Online, 2020: pp. 7203–7219. doi:10.18653/v1/2020.acl-main.645.

[18] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP, in: Proc. 2019 Conf. North, Association for Computational Linguistics, Minneapolis, Minnesota, 2019: pp. 54–59. doi:10.18653/v1/N19-4010.

[19] K. Cho, B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, ArXiv. abs/1406.1078 (2014).

[20] F.A. Gers, and J. Schmidhuber, LSTM recurrent networks learn simple context-free and context-sensitive languages, IEEE Trans. Neural Netw. 12 6 (2001) 1333–40.

Address for correspondence

marc.vincent@institutimagine.org