World Patent Information 49 (2017) A1–A3

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Information retrieval, machine learning, and Natural Language Processing for intellectual property information

Keywords:
Editorial
Patent information
Machine learning
Information retrieval
Natural Language Processing

A B S T R A C T

Readers of this journal are well aware that automation technology has played a significant role in searching for patent information and, as artificial intelligence is once again (after the first, 1960s, and second, 1980s, golden eras of AI) a trending topic at both academic and industry conferences, the editorial team of this journal would like to encourage contributions that cover any aspect of automation as applied to intellectual property information.

As a new Associate Editor of World Patent Information, I take the opportunity to advertise the availability of the editorial team to submissions from computer science teams, in addition to the existing contributions from the IP community.

By way of introduction, I have taken the liberty to provide an extremely brief overview of efforts and contributions that the computer science community has made to the field of intellectual property. This summary focuses on patents, but that is not to say that trademarks or other forms of intellectual property have not triggered the interest of computer scientists. In fact, to call this summary short is already giving it too much credit. It is but a seed, upon which I hope that a forest of contributions and publications from my fellow computer scientists will grow.

1. The past

Looking back at this particular Journal, we see that with a single exception (Volume 44 in March 2016), the past 10 vol had at least one contribution directly addressing the usage of systems or software for exploring, finding, or analysing patent information. We learned about complete systems, like PerFedPat [1] or, more recently, Patent2Net [2]; which presented very specific algorithms addressing new developments in the industry (e.g. detecting obviousness in 3-D printing materials [3]); and, particularly interesting for the technology developers among us, we read a lot about how technology is actually used, what issues are still to be solved and where the major pain points are.

On the other hand, the computer science community has also not been quiet on the topic. A quick search on the DBLP Computer Science Bibliography shows that over the past 10 years there have been 1170 publications with the word “patent” in their title or abstract (see Fig. 1). However, we also see that after a peek in 2010–2014, the number of patent related publications (grossly estimated by the keyword based search) has decreased significantly in 2015 and 2016.

As Associate Editor of World Patent Information, I see here an opportunity for our journal to attract excellent research results and encourage the exchange between different application domains of the technologies developed in the computer science community. Such application domains have and will continue to cover everything from management sciences (e.g. automatic patent valuation [4]) to machine translation [5], not forgetting visualisation [6].

2. The present

Today we are in a position where the presence of the patent domain in computer science venues has decreased, as we have seen in Fig. 1. My personal opinion on the causes of this is that the computer science academic community is lacking a champion. It is much easier to publish (the main recognition in the academic world) on more trending topics (web, social medial, healthcare). I say ‘easier’ not in the sense that these domains do not have complex challenges of their own, but because there is strength in numbers, and each individual author does not have to explain again, within the limited space of a scientific article, the motivation behind his or her research. However, this is not to say that there is no interesting computer science research being applied to the patent domain. We still see highly pertinent outputs on automated patent valuation [7], text summarisation [8], and chemical entity recognition [9].

Above all however, the nature of the stylistic language used in patents has been a constantly recurring topic of discussion both in this journal [10] and elsewhere [11]. Recent progress has been made in the detection of technical terminology, particularly with the purpose of generating useful queries based on full-text patent

http://dx.doi.org/10.1016/j.wpi.2017.06.002
0172-2190/

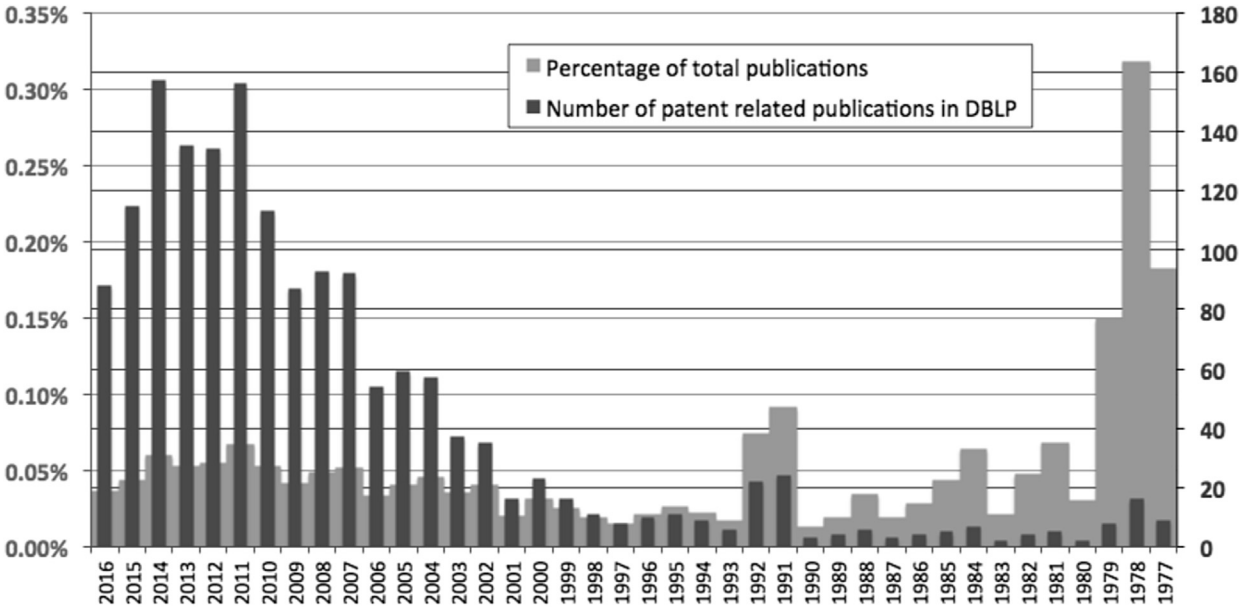

Fig. 1. Number and percentage of patent related publications the DBLP (http://dblp.org/) index of computer science publications.

applications [12]. The past few years have also seen an adoption in the search technologies area of methods and models recently revolutionising machine learning, namely so called *deep learning* methods. While these have been used with extreme success there where the data is signal-like (images recognition works at human levels, as shown in Fig. 2), their application on text has been relatively slow. We see excellent results on short texts, particularly in Natural Language Processing (NLP) tasks such as sentence parsing [13] or sentiment analysis [14]. For ranking based on relevance of the full text of a document to a query, the first workshop on the topic (i.e. the application of neural models in information retrieval) was held in 2016 [15].

Currently, we see modern neural networks primarily applied for machine translation. By modern here we understand essentially larger and with more complex structures than what was available 50, or even 10 years ago. Some open-source code is already available,[1] as a collaborative effort between Systran and the Harvard NLP group, based on neural models developed at Stanford [16] and at Bremen and Montreal [17]. For the rest of the patent-related applications (classification, extraction, IP management), there is still a wide open field of opportunities. Neural networks have been previously applied in the patent domain for other topics than machine translation, such as patent analysis [18], patent valuation [19], and of course patent classification [20], but to the extent to which the recent models of neural networks are different from those of a the previous decade, we are still to see new experiments using these models.

3. The future

What trends can we expect to see for the future in technologies related to intellectual property retrieval and analysis? On the short term, the remarkable success of deep learning in other application domains will almost certainly be tested on patent or other intellectual property data. I hope that the authors will share their results and observations with the community, and that they will consider

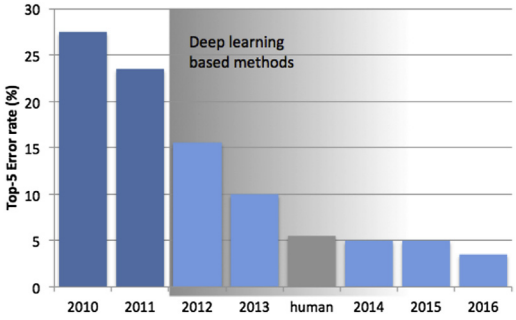

Fig. 2. Error rate in the ImageNet Large Scale Visual Recognition Challenge over the years, showing how deep learning methods have reached human-level performance.

publishing in this journal. Especially as applying them, successfully, to patent data in particular is not immediately obvious. Take for instance the example of images, where deep learning has had the most visible success [21]. In terms of patent drawings however, we do not yet see the same level of success [22], though at the moment of writing I am not aware of any academic report on using neural networks for this task. The lack of features to learn from, the black-and-white nature of patent drawings makes this not particularly surprising.

In the case of text, the situation is different. We are already seeing machine translation results (though an objective and independent evaluation in the style of the NTCIR benchmarks for machine translation is still missing [23]). On other tasks, such as search and information extraction, I look forward to seeing these methods brought to bear on the patent information seeking scenarios familiar to the community of this journal.

4. In this volume

Before we get to the future, let us look at some highlights within the current volume.

We begin with an overview and an update on regions of particular interest for our readers: BRICS countries first, and China second. Deorsola and her colleagues at the Brazilian National

[1] http://opennmt.net.

Institute of Industrial Property studied the different practice among BRICS countries in terms of intellectual property in general and trademarks in particular. Their report shows a very useful table of treaties to which each of these five countries have adhered to, but also go into details of what can and cannot be trademarked in each of them.

To continue your research on BRICS, China, trademarks and the wider topics of intellectual property, Susan Bates has compiled a literature listing with the latest books, journal articles, covering everything from policy making to software tools, while not forgetting historical perspectives on intellectual property.

We start the research section of this issue with a study by Mark James Thomson at the Swiss Federal Institute of Technology, who, using recently collected data on fees across a wide set of patent authorities, builds on the relative lack of studies on renewal and maintenance fees (as opposed to application fees) and observes that patent owners tend to see the value of maintaining their protection higher than the current fees. This, in the opinion of the author, provides support to arguments of increasing maintenance and renewal fees. This would certainly leave some of our readers in disagreement, and we look forward to further studies and publications on this matter here.

Also in support of decision makers, the second research article of this issue, collaboratively written by a diverse team from Germany, USA, and Spain, describes a novel software system, PatStream, to visualise changing topics in patents over time. Their study is a result of a EU Commission funded research project, iPatDoc, and combines not just a visualisation tool, but also, and perhaps more importantly, text analytics tools to populate the data underlying the visualisation. Another very interesting aspect of their data is the so-called *innovation* score of each patent, calculated based on the differences and similarities between the current patent and the previous and future patents. This innovation score is inversely proportional with the similarity of the current patent o previous patents, and directly proportional with its similarity to future patents (i.e. under the intuition that a patent with high impact will shape future research and products).

Carrara and Russo, from the Università degli Studi di Bergamo provide a compendium of all the issues that should be considered when reviewing documents for a risk analysis such as 'freedom to operate'. This review of the changes that can occur to the legal scope of a document post publication will I am sure be useful to those working in industry.

A horizontal study across years and technology domains, but concentrated only on Chile is presented next. Pinto and his colleagues at the Universidad Católica del Norte collected data from the Chilean National Institute of Industrial Property over a span of 25 years from 1989 to 2013 and observed innovation patterns in their country. While some observations match those of other jurisdictions (for instant that Universities have an almost negligible percentage of patents), some are clearly specific to Chile (e.g. the emphasis on the mining industry). It is also interesting that the study covers the transition of Chile from the pre-PCT to the post-PCT era, and this has an apparent effect on the grant lags.

Finally, this issue concludes with a tribute to Eugene Garfield, the conferences diary, as well as a report about a new partnership between the French Patent Information Association (CFIP) and the European Institute for Enterprise and Intellectual Property (IEEPI), signed at the end of May 2016. The new partnership, built around sharing expertise and networks, communication, and training should be of particular relevance to our French-speaking readership, but also to all interested in extending their knowledge on intellectual property in France.

I hope you will enjoy reading this issue as much as I have and look forward to your comments and future contributions to our journal!

References

[1] M. Salampasis, A. Hanbury, PerFedPat: an integrated federated system for patent search, World Pat. Inf. 38 (2014) 4–11.
[2] D. Reymond, L. Quoniam, A new patent processing suite for academic and research purposes, World Pat. Inf. 47 (2016) 40–50.
[3] J.M. Pearce, A novel approach to obviousness: an algorithm for identifying prior art concerning 3-D printing materials, World Pat. Inf. 42 (2015) 13–18.
[4] A.L. Porter, S.W. Cunningham, Tech Mining: Exploiting New Technologies for Competitive Advantage, John Wiley & Sons, 2004.
[5] D. Wang, Chinese to english automatic patent machine translation at {SIPO}, World Pat. Inf. 31 (2009) 137–139.
[6] Y. Yang, L. Akers, T. Klose, C.B. Yang, Text mining and visualization tools – impressions of emerging capabilities, World Pat. Inf. 30 (2008) 280–293.
[7] X. Liu, J. Yan, S. Xiao, X. Wang, H. Zha, S. M. Chu, On predictive patent valuation: forecasting patent citations and their types, Proceedings of the AAAI Conference on Artificial Intelligence, 1438–1444.
[8] J. Codina-Filbà, N. Bouayad-Agha, A. Burga, G. Casamayor, S. Mille, A. Müller, H. Saggion, L. Wanner, Using genre-specific features for patent summaries, Inf. Process. Manag. 53 (2017) 151–174.
[9] M. Habibi, D.L. Wiegandt, F. Schmedding, U. Leser, Recognizing chemicals in patents: a comparative analysis, J. Cheminformatics 8 (2016) 59.
[10] S. van Dulken, Do you know English? The challenge of the English language for patent searchers, World Pat. Inf. (2014).
[11] K. H. Atkinson, Towards a more rational patent search paradigmProc. of PaIR.
[12] L. Andersson, M. Lupu, J. R. M. Palotti, A. Hanbury, A. Rauber, When is the time ripe for natural language processing for patent passage retrieval?Proc. of CIKM.
[13] S. R. Bowman, J. Gauthier, A. Rastogi, R. Gupta, C. D. Manning, C. Potts, A fast unified model for parsing and sentence understandingProceedings of the ACL.
[14] M. Yang, W. Tu, J. Wang, F. Xu, X. Chen, Attention based LSTM for target dependent sentiment classification, in: Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence, February 4-9, 2017, pp. 5013–5014. San Francisco, California, USA.
[15] First International Workshop on Neural information Retrieval, https://www.microsoft.com/en-us/research/event/neuir2016/, 2016.
[16] M.-T. Luong, H. Pham, C. D. Manning, Effective approaches to attention-based neural machine translationProc. of EMNLP.
[17] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translateProc. of ICLR.
[18] J.-C. Lamirel, S. Al Shehabi, M. Hoffmann, C. François, Intelligent patent analysis through the use of a neural network: experiment of multi-viewpoint analysis with the multisom model, Proceedings of the ACL-2003 Workshop on Patent Corpus Processing - Volume 20, 7–23.
[19] Y.-H. Lai, H.-C. Che, Modeling patent legal value by extension neural network, Expert Syst. Appl. 36 (2009) 10520–10528.
[20] A.J. Trappey, F.-C. Hsu, C.V. Trappey, C.-I. Lin, Development of a patent document classification and search platform using a back-propagation network, Expert Syst. Appl. 31 (2006) 755–765.
[21] https://www.dsiac.org/resources/journals/dsiac/winter-2017-volume-4-number-1/real-time-situ-intelligent-video-analytics, 2017.
[22] S. Vrochidis, A. Moumtzidou, I. Kompatsiaris, Enhancing patent search with content-based image retrieval, in: G. Paltoglou, F. Loizides, P. Hansen (Eds.), Professional Search in the Modern World - COST Action IC1002 on Multilingual and Multifaceted Interactive Information Access, Volume 8830 of *Lecture Notes in Computer Science*, Springer, 2014, pp. 250–273.
[23] M. Lupu, A. Fujii, D. Oard, M. Iwayama, N. Kando, Patent-related Tasks at NTCIR, in: Current Challenges in Patent Information Retrieval, second ed., 2017, pp. 77–111.

Mihai Lupu
TU, Wien, Vienna, Austria
E-mail address: mihai.lupu@tuwien.ac.at.