Annals of Library and Information Studies
Vol. 65, March 2018, pp. 62-69

New bibliometric indicators for prospectivity estimation of research fields

Ravil I. Muhamedyev[a,c*], Ramiz M. Aliguliyev [b], Zhassulan M. Shokishalov [c] and Rinat R. Mustakayev[c]

[a]Coresponding author: Kazakh National Research Technical University, Kazakhstan, Almaty, Satpaev 22,
Email: ravil.muhamedyev@gmail.com

[b]Institute of Information Technology, Azerbaijan National Academy of Sciences, Azerbaijan, Baku, AZ1141, B. Vahabzade str. 9A.

[c]Institute of Information and Computational Technologies, Kazakhstan, Almaty, Pushkin str. 125

Received: 24 July 2017; revised: 19 February 2018; accepted: 25 March 2018

The paper suggests differential metrics for estimation of change dynamics of major ICT fields using the bibliometric indicators (publication and citation count). It refers to research areas such as big data, computational biology, cloud computing, cyber-physical systems, embedded systems, information security, internet of things, human-machine systems, mobile computing, machine learning, machine-to-machine, multi-agent systems, neural networks, robotics, visualization, augmented reality, SDN, 5G, e-Governance, smart city and smart grid. As supplements to the known indicators, two kinds of integrated derivative-based indicators are suggested. The calculation of indicators is made and their time curve is given. The suggested indicators allow evidently expressing the changes in the dynamics of bibliometric indicators, which can be useful in prospectivity estimation of areas of research.

Keywords: Bibliometric; Scientometrics; ICT fields; Citations; Differential indicators

Introduction

Information and communication technologies (ICTs) are a rapidly growing research front. In the recent years, new areas such as big data, bioinformatics (computational biology), cloud computing, cyber-physical systems, embedded systems, information security, internet of things, human-machine systems, mobile computing, machine learning, machine-to-machine, multi-agent systems, neural networks, robotics, visualization, augmented reality, SDN, 5G, e-Governance, smart city, smart grid etc., have come into being. Table 1 gives the annual publication count from Elsevier's ScienceDirect in each of the mentioned areas.

Analysis of ICT research areas above by means of publication count in the mentioned subfields allows the assessment of their growth dynamics. Identification of patterns by publication count gives only the most common estimation of prospectivity for one area of research or another. It is important to assess relative size of assessed research fields, their interrelation and change dynamics using scientometric indicators.

Some earlier studies have formulated the goals of the research and have suggested some formal characterization[1,2,3]. A semantic network has been suggested for the estimation of research field interrelation. Previous studies have also introduced the concepts of similarity and constraint forces between the concepts, which can be used for further formal characterization of a semantic network[4,5,6]. However, formal estimation of change dynamics of scientometric indicators in ICT filed has not been performed.

This article introduces formalized metrics of estimation of change dynamics of scientometric indicators (indicators of "prospectivity") that allows making an integrated evaluation of one area of research or another. We used the set of ICT fields considered in previous studies to verify the workability of introduced indicators[1,3].

Review of Literature

The idea of unified publications assessment in the field of science by applying the citation index (SCI)

Table 1—Change dynamics of publication count in ICT fields by ScienceDirect data

Field	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Augmented Reality	1137	1355	1379	1494	1459	1571	1779	1984	2348	2453	2790	2458
Big Data	18501	21216	23489	25098	26886	27521	31395	34400	38617	43535	50701	48974
Bioinformatics	3337	3800	4201	4598	5127	5483	5740	6915	7714	8613	9731	9065
Cloud computing	1663	1818	2000	2098	2369	2581	3463	4007	5266	5869	7118	6940
Cyber-Physical systems	194	187	270	229	234	267	486	479	796	1040	1290	1284
Embedded systems	23671	27234	28237	29803	37599	32764	35655	38545	42736	45403	50469	46875
Information Security	8740	9649	11234	11184	11717	12479	14005	15549	17871	20150	22557	21208
Internet of things	2945	3051	3953	3689	3481	3909	3967	4322	4567	5150	6276	5622
Human-machine systems	6280	7349	7689	8027	8571	8904	9798	11227	12659	13750	16188	15457
Mobile computing	3523	4009	4566	4867	4967	5088	6049	6642	7724	8491	9753	8883
Machine Learning	5025	5914	6355	6824	7291	7378	8500	9566	10720	12104	14777	14242
Machine to machine	24644	28150	28930	30158	32302	31884	36139	39279	43918	47746	54137	51479
Multi agent systems	10098	10843	11897	12885	13472	14260	15766	17690	20006	21417	23704	22605
Neural Networks	8876	10088	10255	11442	13153	11910	13625	14572	15598	16844	19012	17939
Robotics	3191	3727	3702	4783	5120	5121	6080	6806	7029	7894	8418	7377
Visualization	14030	15113	15627	16751	17980	18384	20007	21969	23441	25114	27892	26197
Intelligent transport systems	835	924	1050	1073	1107	1153	1403	1641	2138	2176	2466	2393
E-Governance	1809	2227	2321	2738	3116	3327	3727	4526	5540	6651	7653	6788
Software Defined Networking	9059	10949	11787	12727	14313	14869	16899	19123	22089	24633	27278	26434
5G	8413	10425	11509	12795	19782	15093	16301	17067	18982	20160	21585	20239
Smart city	821	935	1093	1132	1195	1352	1547	1917	2179	2748	3540	3846
Smart grid	447	505	619	694	850	1077	1588	1892	2366	3075	4074	4023

was proposed by Eugene Garfield[7]. Bibliometric indicators, such as the number of publications, the citation index, the number of co-authors, etc., are widely used for the evaluation of research fields and organizations in selected countries and regions[8-11], policy making in the field of scientific research[12], impact assessment of publications databases, in particular, the use of public archives increases citation of publications[13]. Some authors use a bibliometric data to build prediction models[14], in some cases in combination with patent analysis[15,16] and to assess the development of science-intensive areas (https://clarivate.com/stateofinnovation).

At the same time, bibliometric methods have significant limitations. In particular, numerical indexes are non-linearly dependent on the size of the country and organization[17]. For example, the larger the research group, the higher the Crown indicator and h-index[18]. In addition, h-index is very sensitive to research fields characteristics (productivity, citation habits, and citation dynamics). The use of indicators without a clear understanding of subject area leads to the effects of "quick and dirty"[19]. In this regard, various authors introduce new or modify existing indicators. For example, the change in the Hirsch index with time has been estimated[20]. The special index (j-index) was introduced[21]. The j-index is a square root of the numbers of citations to each of the author's publications. It has been reported that the new index better describes the field of scientific publications of the author[21].

To assess the dynamics of the development of scientific areas, it is necessary to assess the velocity of change. This need was realized by analyzing the temporary graphs of indicators. In spite of the fact that the h-velocity index introduced to estimate the rate of change of Hirsch index, the numerical estimation of changes in bibliometric indicators is not widely implemented[22]. In our opinion, the use of dynamic indicators along with full-text analysis allows us to assess more accurately the potentials of research areas[23].

Objectives of the study

- To introduce new bibliometric indicators (differential indicators) to estimate dynamic parameters of publication activity in ICT field;
- To apply these indicators to estimate dynamics of ICT subfields; and

- To estimate ICT fields dynamics using differential indicators.

Prospectivity metrics of research fields

The estimation of the change dynamics of publication activity, and alongside the "prospectivity" of definite fields of research, requires certain formal characterization.

Informally, prospectivity of one scientific field or another reflects its relevance now or in the future. We consider prospectivity as a measure for achievement of significant scientific results. We can say that the prospectivity is a function of relevance: Prospectivity = f (Relevance).

The concept of relevance, same as the prospectivity, is difficult to be exactly defined. Relevance depends on economic, social, psychological and other factors. Assessment of relevance was made by the authors of publications, who perform scientific research in one area or another. Authors' interest becomes evident through publication and citation count in the given area of research. Relevance also shows up in change dynamics of scientific publication and citation count. If the change dynamics is not considered herein, the areas that have worked out themselves from the point of new scientific results could be accepted as prospective ones. In the estimation of this dynamics there can be applied Compound Annual Growth Rate – CAGR, which is calculated according to the following Eq. (1):

$$\text{CAGR} = \left(\frac{\text{Ending Value}}{\text{Beginning Value}}\right)^{\frac{1}{\text{T}-1}} - 1 \quad \ldots(1)$$

where T is the number of periods. It is clear that the scientific areas demonstrating growth of this rate by publication and citation count can be classified as prospective ones.

Nevertheless, as outlined below, this rate demonstrates that rather smooth change, hindering identification of change dynamics patterns, is calculated individually for publications and for citations. In this regard, it seems to be useful to introduce integrated indicators for estimation of change dynamics. For this purpose, we will determine the following differential indicators of a prospective scientific area, defined by the given key search term

$$D1_i = f_1(n_i, \frac{dn_i}{dt}, \frac{dc_i}{dt}) \quad \ldots(2)$$

$$D2_i = f_2(n_i, \frac{d^2 n_i}{dt^2}, \frac{d^2 c_i}{dt^2}) \quad \ldots(3)$$

Put it differently, prospectivity indicator $D1_i$ is a function depending on the publication count n_i, speed of change of the publication count $\frac{dn_i}{dt}$ (time derivative) and the citation count $\frac{dc_i}{dt}$ in the given area of research.

Indicator $D2_i$ is a function depending on the publication count n_i, acceleration of change of the publication count $\frac{d^2 n_i}{dt}$ and the citation count $\frac{d^2 c_i}{dt}$ in the given area of research.

Functions, f_1 and f_2, aggregate contribution of citations and publications by some means or other. In a particular case, aggregation can be accomplished using weighted summation. Then, for an individual scientometric database j, scientific research prospectivity indicators in a field i at the moment of time t_k, can be determined with the Eqs. (4) and (5)

$$D1_i^j(t_k) = \alpha \times n_i^j(t_k) + \beta \times \frac{dn_i^j(t_k)}{dt} + \gamma \times \frac{dc_i^j(t_k)}{dt} \quad \ldots(4)$$

$$D2_i^j(t_k) = \alpha' \times n_i^j(t_k) + \beta' \times \frac{d^2 n_i^j(t_k)}{dt} + \gamma' \times \frac{d^2 c_i^j(t_k)}{dt} \quad \ldots(5)$$

where $\alpha, \beta, \gamma, \alpha', \beta', \gamma'$ are some empirical factors to be regulating contribution share of publication count, speed and acceleration of change of the publication count n_i and speed and acceleration of change of the citation count c_i, respectively. Note that if the second derivative function can be calculated using the first derivative $\frac{d^2 f}{dx^2} = \frac{d(df/dx)}{dx}$, so the Eq. (5) can be expressed as follows:

$$D2_i^j(t_k) = \alpha' \times n_i^j(t_k) + \beta' \times \frac{d(dn_i^j(t_k)/dt)}{dt} + \gamma' \times \frac{d(dc_i^j(t_k)/dt)}{dt} \quad \text{...(6)}$$

For more evident estimation of change dynamics in publishing activity, especially in the fields with large amount of publications, indicators α and α' can be made equal to 0. In this case

$$D1_i^j(t_k) = \beta \times \frac{dn_i^j(t_k)}{dt} + \gamma \times \frac{dc_i^j(t_k)}{dt} \quad \text{...(7)}$$

$$D2_i^j(t_k) = \beta' \times \frac{d(dn_i^j(t_k)/dt)}{dt} + \gamma' \times \frac{d(dc_i^j(t_k)/dt)}{dt} \quad \text{...(8)}$$

Eqs. (1), (7) and (8) were used for further calculations.

"Prospectivity" of ICT Fields

Tables 1 and 2 show the search results. The year 2016 data is valid until the mid-year. Figures 1 and 2 are obtained according to the data from Tables 1 and 2.

The result was obtained by a simple method. We calculated the number of articles and citations containing the relevant keywords (augmented reality, big data, etc.) by years, since 2005. We used keywords related to ICT field that describes some relatively new fields of research[1,3].

Derivative is calculated using the equation $f'(x0) \approx (f(x0+\Delta x) - f(x0-\Delta x))/2\Delta x$, where x0 – is a current year, $\Delta x = 1$ (a minimum period of data acquisition equal to 1 year). Due to specifics of derivative calculations, D1 indicator can be calculated for the period from 2006 till 2014, but D2 – from 2007 till 2013. By way of illustration, Tables 3 and 4 present results of CAGR, D1, D2 calculations in the field machine learning. Calculation results for the rest of fields are shown as curves.

The curves in Figure 3 illustrate change dynamics of D1 and D2 in ICT fields. As D1 effectively means speed of change, but D2 – acceleration of publication and citation count growth, they are indicated in the curves as Speed and Acceleration, respectively.

Table 2—Change dynamics of citation count in ICT fields by ScienceDirect data

Field	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Augmented Reality	119	116	134	161	174	229	254	313	442	445	611	615
Big Data	665	817	933	1076	1191	1354	1554	1857	2396	2984	4137	5007
Bioinformatics	2414	3126	3649	4220	4721	5219	5745	6720	7976	8320	9590	9271
Cloud computing	95	107	139	147	166	248	478	696	1091	1464	2105	2433
Cyber-Physical systems	11	15	29	31	21	37	73	107	73	297	456	552
Embedded systems	1417	1714	2078	2323	2808	2996	3510	4169	5035	5667	6989	7682
Information Security	1121	1252	1711	1702	2121	2282	2849	3438	4082	4565	5078	5446
Internet of things	366	396	511	488	559	554	728	921	1138	1425	1506	1759
Human-machine systems	1278	1486	1677	1957	2201	2328	2709	3161	3671	3897	4559	4675
Mobile computing	621	798	1070	1299	1377	1479	2041	2394	2708	2904	3582	3596
Machine Learning	1954	2538	2899	3530	4220	4176	5205	5914	6703	7489	8982	9301
Machine to machine	6741	7910	8465	9688	10783	10633	12569	14304	15925	17111	19747	20082
Multi agent systems	1363	1680	1848	2243	2509	2768	3380	4097	4807	5741	6732	7138
Neural Networks	5911	6930	7797	9120	10360	10163	12384	13590	15455	16691	19363	20026
Robotics	1557	1822	2015	2595	2851	2962	3522	4150	4488	4852	5364	5045
Visualization	3862	4155	4568	5271	5501	5906	6534	7074	7906	8558	9855	9952
Intelligent transport systems	137	216	230	286	343	397	550	761	931	1085	1207	1209
E-Governance	872	1150	1340	1685	1988	2244	2650	3293	4235	4955	5971	5961
Software Defined Networking	92	163	196	220	283	311	354	500	643	721	858	967
5G	69	86	78	89	113	93	100	134	133	134	170	240
Smart city	127	168	199	222	214	278	346	472	633	878	1296	1590
Smart grid	29	48	54	77	95	145	298	548	803	1458	2062	2577

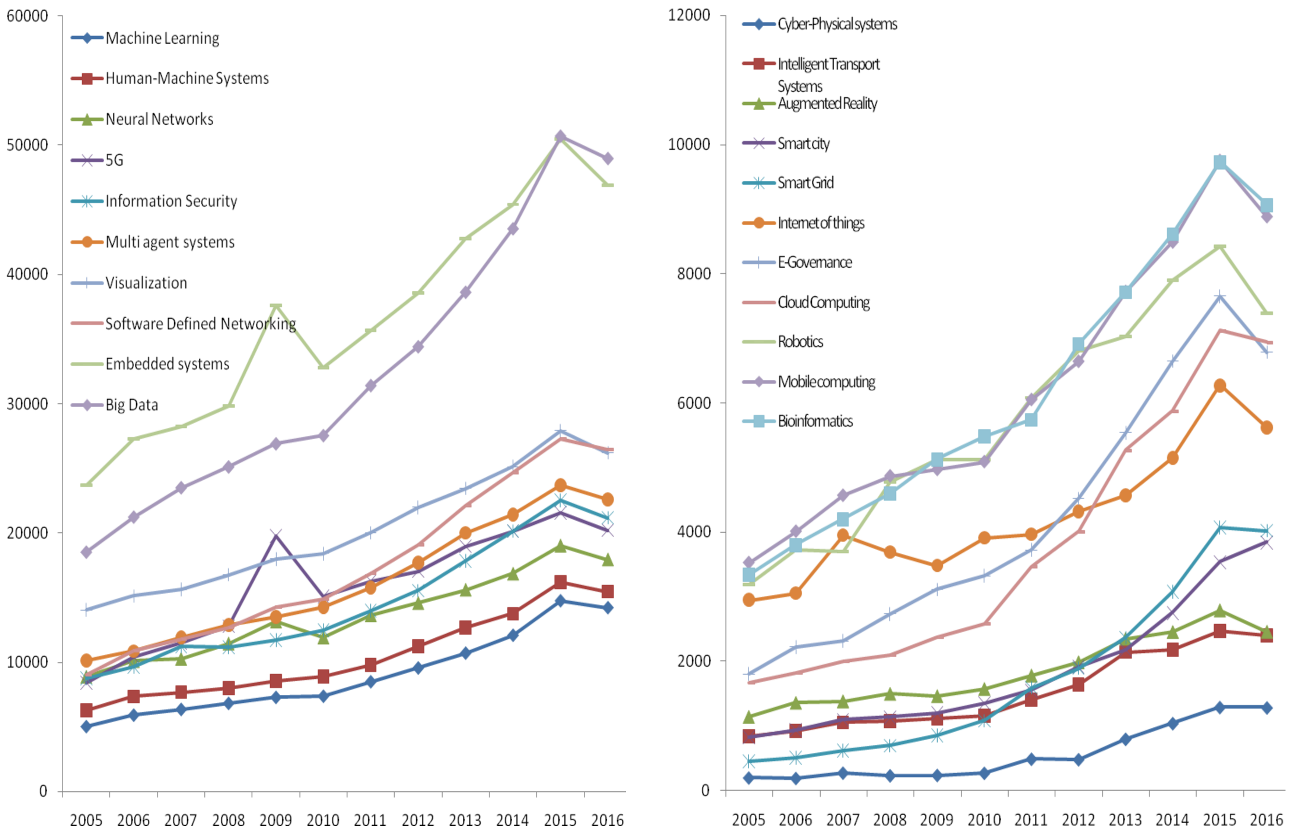

Fig. 1—Change of publication count in ICT fields by ScienceDirect data

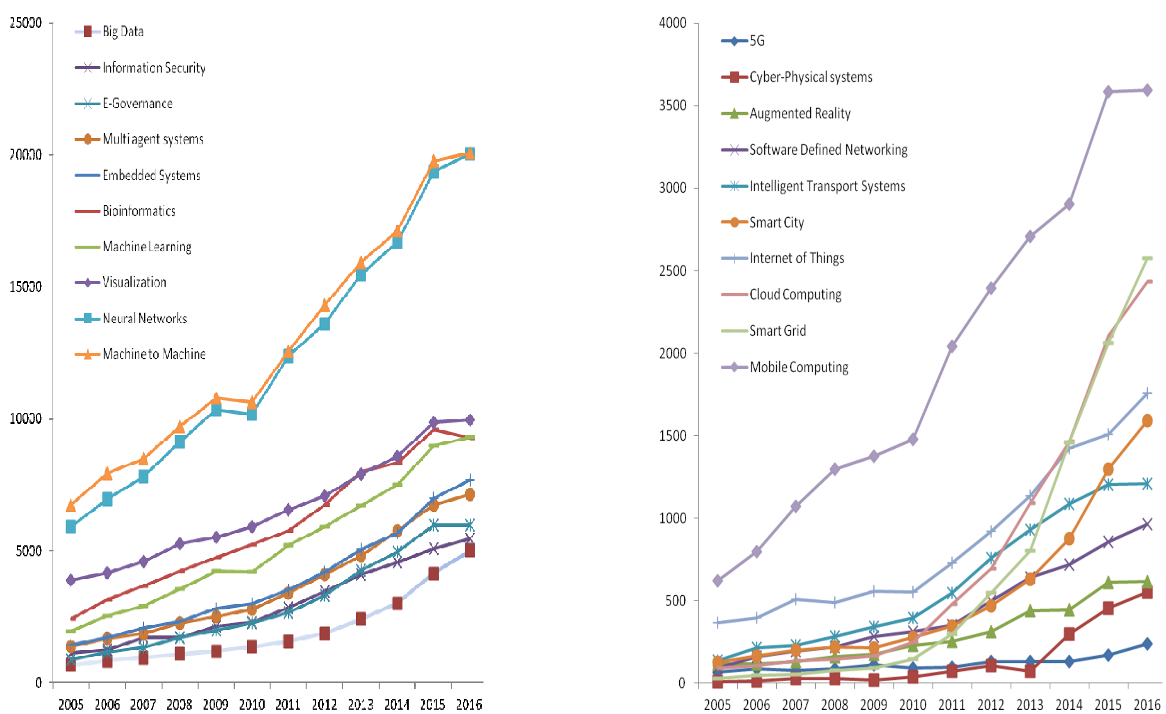

Fig. 2—Change of citation count in ICT fields by ScienceDirect data

Table 3—Initial data and calculation results for CAGR indicators in the Machine Learning field

Year	Publication count (*n*)	T	Citation count (c)	CAGR(citations)	CAGR(Publications)
2005	5025	2	1954		
2006	5914	3	2538	13.97	8.49
2007	6355	4	2899	14.05	8.14
2008	6824	5	3530	15.93	7.95
2009	7291	6	4220	16.65	7.73
2010	7378	7	4176	13.49	6.62
2011	8500	8	5205	15.02	7.80
2012	9566	9	5914	14.85	8.38
2013	10720	10	6703	14.68	8.78
2014	12104	11	7489	14.38	9.19
2015	14777	12	8982	14.87	10.30

Table 4—Calculation results for D1 and D2 indicators by the field Machine Learning data

Year	dn/dt	dc/dt	D1 (Speed)	D2 (Acceleration)
2005				
2006	665.00	472.50	98.10	
2007	455.00	496.00	83.50	–50.60
2008	468.00	660.50	100.20	–97.20
2009	277.00	323.00	52.80	29.40
2010	604.50	492.50	95.10	408.70
2011	1094.00	869.00	170.10	240.70
2012	1110.00	749.00	160.00	57.80
2013	1269.00	787.50	176.30	426.00
2014	2028.50	1139.50	270.50	
2015				

Conclusions

Currently, publication count in all the considered semantic ICT sectors demonstrates a steady growth. In this regard we suggest differential indicators of estimation of D1 and D2 prospectivity alongside with CARG indicator for quantitative analysis of the given growth dynamics.

However, if D1 refers to growth rate of scientific publications, application of D2 lets identify growth rate diminution periods of scientific publication and citation count. Negative change of the indicator reveals exactly diminution of growth rate. The data reveals an interesting pattern for majority of new fields, when the period of initial growth is followed by diminution and then - by repeated acceleration (big data, augmented reality, cyber-physical systems, information security, internet of things, human-machine systems, mobile computing, machine learning, machine-to-machine, multi-agent systems, intelligent transport systems, software defined networking). The reasons of such a phenomenon should be investigated but it can be assumed that the specified dynamics characterize the development intensity in the area of research, and on the other hand the comprehension of a new concept by researchers and its application in research. Only two of the presented fields (smart grid, cloud computing) feature a constant steady growth of citation and publication count (D2 indicator has only positive values during the whole period under consideration).

Some fields (neural networks, robotics, visualization) are characterized by repeated acceleration (D2 indicator that had a negative value for several periods becomes positive again), whereas

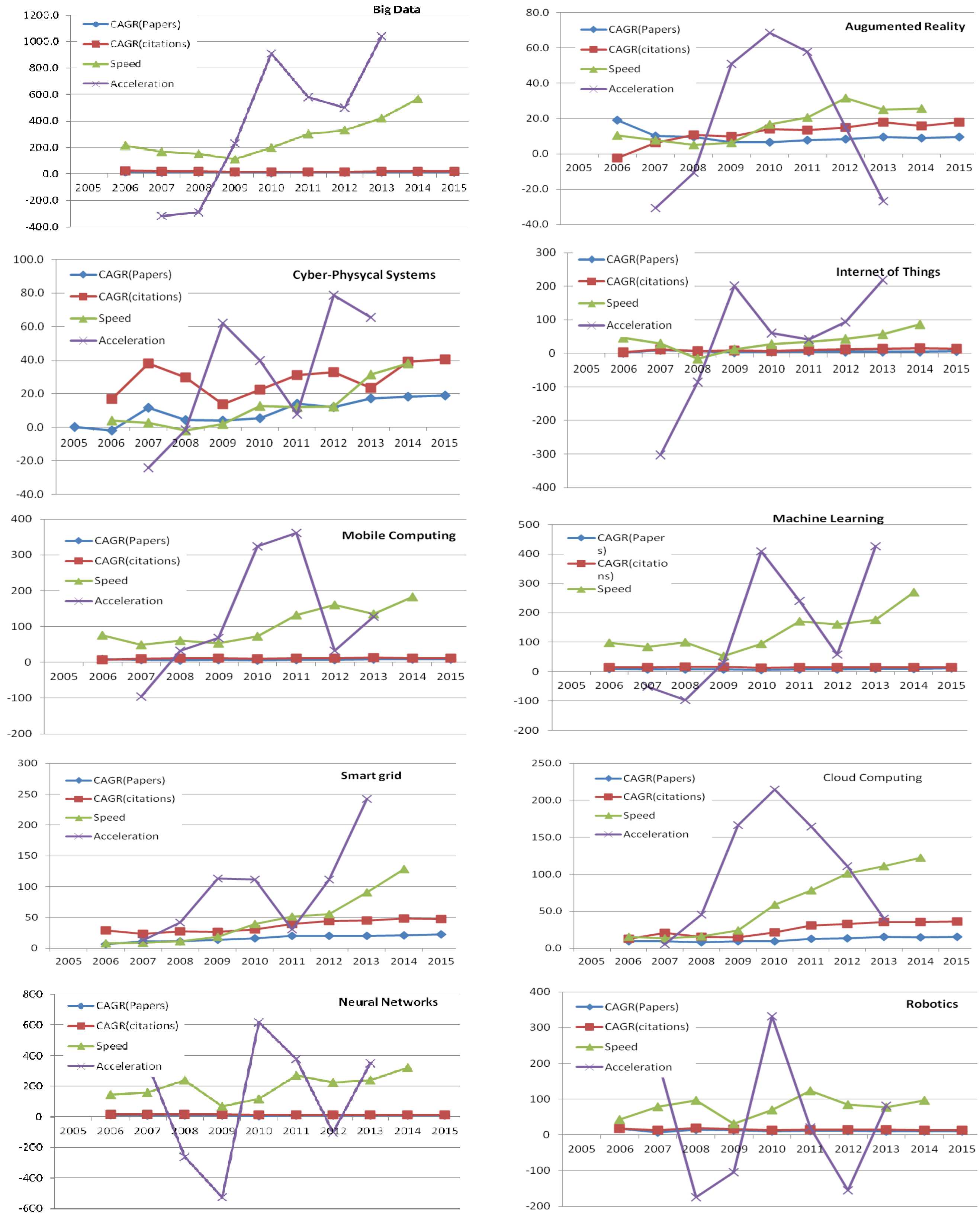

Fig. 3—CAGR, D1 (speed) and D2 (acceleration) indicators of ICT fields

for the fields, smart city and e-governance, the drop period is the same.

It is obvious that the indicators under consideration cannot fully characterize prospectivity of one or other areas of research. When identifying science development vectors at macro level, scientometric approaches alone are not sufficient and there is needed an expert opinion[24]. However, the introduced indicators, in view of their visibility, can be used to support expert decision-making.

Acknowledgements

The research work was completed under the partial support with the Grant 0168 of GF4 Ministry of Education and Science of the Republic of Kazakhstan. Work by Ramiz Aliguliyev was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – Grant № EIF-2014-9(24)-KETPL-14/02/1.

References

1. Muhamedyev R I, Kalimoldaev M N and Uskenbayeva R K, Semantic network of ICT domains and applications, *In Proceedings of the 2014 Conference on Electronic Governance and Open Society: Challenges in Eurasia*, St. Petersburg, Russian Federation, November 18-20, 2014, p. 178-186.
2. Muhamedyev R I, Muhamedyeva J, Amirgaliyev Y N, Khamitov A N and Abdilmanova A, Revelation of new ICT domains for upcoming Kazakhstan's participation, *In Proceedings of the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia,* St. Petersburg, Russian Federation, November 24-25, 2015, p. 179-188.
3. Muhamedyev R I, Amirgaliyev Y N, Kalimoldayev M N, Khamitov A N and Abdilmanova A, Selection of the most prominent lines of research in ICT domain, *In Proceedings of the 2015 Twelve International Conference on Electronics Computer and Computation*, Almaty, Kazakhstan, 27-30 September, 2015, p. 36-42.
4. Aliguliyev R M and Aliguliyev R M, Evolutionary algorithm for extractive text summarization, *Intelligent Information Management*, 1 (1) (2009) 128-138.
5. Cilibrasi R L and Vitányi P M B, The Google similarity distance, *IEEE Transaction on Knowledge and Data Engineering*, 19 (3) (2007) 370–383.
6. Aliguliyev R M, A new sentence similarity measure and sentence based extractive technique for automatic text summarization, *Expert Systems with Applications*, 36 (4) (2009) 7764–7772.
7. Garfield E, Citation indexes for science: a new dimension in documentation through association of ideas, *Science,* 122 (1955) 108–111.
8. Gauthier É, Bibliometric analysis of scientific and technological research: a user's guide to the methodology (Science and Technology Redesign Project, Statistics Canada), 1998, 81 p.
9. Van Raan A F J, The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technology Assessment-Theory and Practice,* 1 (12) (2003) 20-29.
10. Giovanni A, D'Angelo C A and Pugini F, The measurement of Italian universities' research productivity by a non-parametric bibliometric methodology, *Scientometrics*, 76 (2) (2008) 225-244.
11. Mokhnacheva J V and Mitroshin I A, Nanoscience and nanotechnologies at the Moscow domain: a bibliometric analysis based on Web of Science (Thomson Reuters), *Information Resources of Russia*, 6 (2014) 17-23.
12. Koenraad D and Glänzel W, Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key, *Scientometrics*, 59 (2) (2004) 253-276.
13. Moed H F, The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section, *Journal of the Association for Information Science and Technology*, 58 (13) (2007) 2047-2054.
14. Daim T U, Rueda G R and Martin H T, Technology forecasting using bibliometric analysis and system dynamics, *In Technology Management: A Unifying Discipline for Melting the Boundaries*, 31 July-4 Aug 2005, p. 112-122
15. Daim T U, Rueda G, Martin H and Gerdsri P, Forecasting emerging technologies: Use of bibliometrics and patent analysis, *Technological Forecasting and Social Change*, 73 (8) (2006) 981-1012.
16. Takashi I and Squicciarini M, ICT: A new taxonomy based on the international patent classification. *OECD Science, Technology and Industry Working Papers*, 2017/01, OECD Publishing, Paris. http://dx.doi.org/10.1787/ab16c396-en.
17. Sylvan K J, Scale-independent indicators and research evaluation, *Science and Public Policy*, 27 (1) (2000) 23-36.
18. Valérie D and Gevenois P A, Bibliometric indicators: quality measurements of scientific publication, *Radiology*, 255 (2) (2010) 342-351.
19. van Raan A F J, Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods, *Scientometrics,* 62 (1) (2005) 133-143.
20. Egghe L, Dynamic h-index: The Hirsch index in function of time, *Journal of the Association for Information Science and Technology*, 58 (3) (2007) 452-454.
21. Mark L, Fenner T and Bar-Ilan J, A bibliometric index based on the complete list of cited publications, *arXiv*:*1304.6945v1* (2013) 1-12.
22. Rousseau R and Ye F Y, A proposal for a dynamic h-type index, *Journal of the Association for Information Science and Technology*, 59 (11) (2008) 1853-1855.
23. Shvets A, Devyatkin D, Sochenkov I, Tikhomirov I, Popov K and Yarygin K, Detection of current research directions based on full-text clustering, *In Proceedings of the Science and Information Conference*, London, UK, July 28-30, 2015, p. 483-488.
24. Akoev M, Markusova V A, Moskaleva O V and Pislyakov V V, Scientometrics Guide: Indicators of the Development of Science and Technology, (Yekaterinburg, Ural University), 2014, 250 p. (in Russian).