Patent data - The performance and mobility of researchers

Patent information can be used to track the career and performance of individual inventors and to analyse networks of inventors. This information can be used to investigate issues such as researcher mobility (across companies or countries), differences in researcher profiles across fields, linkages across researchers and others, particularly if matched to complementary data. Such use of patent information involves, however, a great deal of data cleaning, as identifying the same individuals in databases with millions of names is not straightforward.

Characteristics of patent data for the analysis of researchers' performance, networks and mobility

The proper identification of inventors in patent filings makes it possible to reconstruct the inventive record of the concerned individuals particularly if it is possible to match this record with complementary data on these individuals available from other databases.

Advances in this area have been hindered by the difficulties associated with the recording of names in patent data and the difficulty of recognizing "who is who" in the population of inventors contained in patent data. Three fundamental problems have made the information on inventors relatively ineffective for investigation. First, the name of the same inventor can be spelled slightly differently across different patents (it may be with or without the middle name or initial, with or without surname modifiers, etc). Second, even if there are two exact same names, it is not certain that the two names correspond to the same person (the "John Smith" problem). In other words, different inventors having exactly the same name may appear in various patents. Third, the transcription into the Latin alphabet of nonwestern names is imperfect and can create ambiguities ("Li" vs. "Lee").

However, researchers have attempted to harmonise names using computerized matching algorithms which they have so far applied to specific subsets of patent data. For example, the methodology developed by Trajtenberg, Shiff and Melamed (2006), which has been used on USPTO patent data, can be summarised as follows:

- Stage 1: group similar names. In order to address the problem of the name of the same inventor being spelled slightly differently from patent to patent, a two-track approach is used. The first is to "clean up" and standardise the names as much as possible; the second is to complete the list of harmonised names, which can be done with the aid of the "Soundex system" to encode names with similar pronunciation. Soundex is a phonetic algorithm for indexing names by sound as pronounced in English. The goal is for names with the same pronunciation to be encoded to the same representation so that they can be matched despite minor differences in spelling.
- Stage 2: compare and match names. To deal with the problem of identifying a given individual among the "suspects" with the same name, the names are compared and matching criteria are imposed. Pair-wise comparisons can be made between any two "suspects" using a series of variables such as middle name, geographic location (e.g. postal codes, cities, etc.), the technological area (i.e. patent class), the assignee, the identity of the co-inventors, etc. If a data item is the same in two suspect records (i.e. if two records display the same address, or are in the same patent class, or share the same partners, etc.), then the pair is assigned a certain score. If the sum of these scores is above a predetermined threshold, the two records are "matched" and they are regarded as being the same inventor. Once that is done for all the pairs in the comparison set, the condition of transitivity is imposed, i.e. if record A is matched to record B, and B to C, then the three are regarded as the same inventor.

Related policy questions that can be addressed by patent data

Policy questions

Patent information can be used to track the career and performance of individual inventors (e.g. their field of work, location, employer), or to analyse networks of inventors (who invents with whom, etc.).

Patents by inventors

A wide array of interesting and highly policy-relevant topics can be investigated with the aid of data on the harmonised names of inventors, including the following:

- The productivity of inventors—over time, across fields, countries, etc. (Hoisl, 2007).

- The mobility of inventors—across cities, regions, countries, sectors (i.e. shifts between the public and private sectors), and the resulting spillovers of such turnover (Kim et al., 2005; Crespi et al., 2005).

- The networking strategies of inventors—who invents with whom—and their impact on productivity (Singh, 2003; Breschi and Lissoni, 2003).

- Gender issues: Share of women and their research profiles (Naldi et al., 2004).

References

- Breschi, S. and F. Lissoni (2003), "Mobility and social networks: Localised knowledge spillovers revisited", CESPRI Working Papers 142, Centre for Research on Innovation and Internationalisation, Universita Bocconi, Milan, Italy

- Crespi, G. A., A. Geuna and L. J. Nesta (2005), "Labour mobility of academic inventors: Career decision and knowledge Transfer", SPRU Electronic Working Paper Series 139, University of Sussex, SPRU—Science and Technology Policy Research, Sussex, UK.

- Hoisl, K. (2007), "Tracing mobile inventors: The causality between inventor mobility and inventor productivity", Research Policy, Vol. 36, pp. 619–36.

- Kim, J., S. J. Lee and G. Marschke (2005), "The influence of university research on industrial innovation", NBER Working Papers 11447, National Bureau of Economic Research, Cambridge, MA.

- Naldi, F., D. Luzi, A. Valente and I.V. Parenti (2004), "Scientific and technological performance by gender", in H.F. Moed et al. (eds.), Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies on R&D Systems, Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 299–314.

- OECD (2009a), OECD Patent Statistics Manual, 5.7, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264056442-en [1].

- OECD (2009b), "OECD Science, Technology and Industry Scoreboard 2009", OECD, Paris. http://dx.doi.org/ [2] 10.1787/sti_scoreboard-2009-en

- Trajtenberg, M., G. Shiff and R. Melamed (2006), "The 'Names Game': Harnessing inventors' patent data for economic research", NBER Working Papers 12479, National Bureau of Economic Research, Cambridge, MA.

Source URL: https://www.innovationpolicyplatform.org/content/patent-data-performance-and-mobility-researchers?topic-filters=12091

Links
[1] http://dx.doi.org/10.1787/9789264056442-en
[2] http://dx.doi.org/