Fair Use in Data Mining and Machine Learning
A Comparative Study between Mainland China and Taiwan
2019 IPSC
Assistant Professor Gi-Kuen Jacob Li
National Tsing Hua University, Taiwan
Institute of Law for Science and Technology
Outline

• Background
  o Technology
  o Legal

• Fair use in data mining & machine learning
  o US
  o Cross-strait comparison
Data mining is the process of “finding valuable patterns in data.”

**SHOLOM M. WEISS ET AL., FUNDAMENTALS OF PREDICTIVE TEXT MINING 1 (2nd ed. 2015).**
Rough Process of Data Mining

Document / Data Collection → Retrieve and preprocess document → Data Analytics → Software / System → Output

access / reproduce (copy)

Reproduce / prepare derivative works (?)

“Expression” / prepare derivative works (?)
Text Mining, Data Mining, and Machine Learning - Techniques and Applications

- Association rule
- Categorization / classification
- Neural network
- Decision tree
- Information retrieval

- Clustering and organizing
- Information extraction
- Prediction and evaluation


Legal Background

• Current US fair use doctrine could adept to resolve latent ambiguities, but there are still questions unanswered

• Fair use for unauthorized, expressive uses of copyrighted works in commercial machine learning are undesirable
Legal Background

• What about jurisdictions that adopt different fair use rules?
  ○ Mainland China followed the Berne three-step test with enumerated limitations
  ○ Taiwan blended categorical illustrations of limitations and US fair use factors into the Copyright Act
Fair Use for Data Mining and Machine Learning

General Issues and Background
“Expressive” and “Non-expressive” Use

- Foundation of Copyright Law
  - Idea-expression dichotomy

- Berne Convention Article 2 (1)
  - “The expression ‘literary and artistic works’ shall include every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression, . . .”

- TRIPS Agreement Article 9 (2)
  - “Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.”
“Copying” in Data Mining and Machine Learning

• Can be “non-expressive”
  o Data mining in healthcare
  o Machine learning for precipitation forecast
  o Crime prediction

“Copying” in Data Mining and Machine Learning

• Copying lead to “expressive” results?

“The technology, which the researchers simply call the drawing bot, can generate images of everything from ordinary pastoral scenes, such as grazing livestock, to the absurd, such as a floating double-decker bus. Each image contains details that are absent from the text descriptions, indicating that this artificial intelligence contains an artificial imagination.”
“Copying” in Data Mining and Machine Learning

• Copying lead to “expressive” results?
“Copying” in Data Mining and Machine Learning

• Copying lead to “expressive” results?
“Copying” in Data Mining and Machine Learning

• Copying lead to “expressive” results?
  - Computoser
    - http://computoser.com/
  - Flow Machines
    - http://www.flow-machines.com/
Walking the Fine Line?

- How to differentiate “expressive” and “non-expressive” use
  - Research results can be utilized in very different way
Not Just A Copyright Issue

• Ethics guidelines for developing AI

• Breach of contract? (terms of use)
  o Binding or not
  o Scope

• Privacy
Fair Use for Data Mining and Machine Learning

Cross-strait Comparison
Limitations of Exclusive Rights (China)

• Article 22: Enumerated Fair Use
  o In the following cases, a work may be used without permission from, and without payment of remuneration to, the copyright owner, provided that the name of the author and the title of the work shall be indicated and the other rights enjoyed by the copyright owner by virtue of this Law shall not be prejudiced:
Limitations of Exclusive Rights (China)

• Article 22: Enumerated Fair Use (cont.)
  (1) use of a published work of others for the user’s own private study, research or self entertainment;
  (2) appropriate quotation from a published work of others in one’s work for the purposes of introduction to, or comment on, a work, or demonstration of a point;
  (3) use of a published work in newspapers, periodicals, radio programs, television programs or newsreels for the purpose of reporting current affairs;
  (4) reprinting by newspapers or periodicals, or rebroadcasting by radio stations or television stations, of editorials or commentator’s articles published by other newspapers, periodicals, radio stations or television stations;
  (5) publication in newspapers or periodicals, or broadcasting by radio stations or television stations, of a speech delivered at a public gathering, except where the author has declared that the publication or broadcasting is not permitted;
  (6) translation or reproduction in a small quantity of copies, of a published work for use by teachers or scientific researchers, in classroom teaching or scientific research, provided that the translation or reproduction shall not be published or distributed;
Limitations of Exclusive Rights (China)

• Article 22: Enumerated Fair Use (cont.)
  (7) use of a published *work by a state organ* for the purpose of performing its official duties;
  (8) reproduction of a work in its collections by a library, an archives center, a memorial hall, a
  museum, an art gallery or a similar institution, for the purposes of *display*, or *preservation* of a
  copy, of the work;
  (9) *free performance* of a published work;
  (10) copying, drawing, photographing, or video recording of an *artistic work* located or on display
  in an *outdoor public place*;
  (11) *translation* of a published work from the language of the Han nationality *into minority
  nationality languages* for publication and distribution in the country;
  (12) *transliteration* of a published work into *Braille* and publication of the work so transliterated.

The above limitations on rights shall be applicable also to the rights of publishers, performers,
producers of sound recordings and video recordings, radio stations and television stations.
Limitations of Exclusive Rights (China)

• Article 23:
Anyone who compiles or publishes **textbooks for the purpose of implementing the nine-year compulsory education or State education planning** may, without the permission from the copyright owner, except that the author has declared in advance that the exploitation is not permitted, compile published fragments of works, short written works or musical works, a single work of fine art, or photographic works into the textbooks, however, he shall pay the remuneration as provided, mention the name of the author and the title of the work, and shall not infringe upon other rights which the copyright owner shall enjoy in accordance with this Law.

The provisions in the preceding paragraph shall be applicable to the limitations on the rights of publishers, performers, producers of sound recordings and video recordings, radio stations and television stations.
Are the Rules Applicable to DM or ML?

- Article 22: Enumerated Fair Use
  - (1) use of a published work of others for the user’s own private study, research or self entertainment;

- (2017) Zhe 0381 Mingchu 832 Hao civil decision:
  - Defendant “linmo” (learn to draw or write calligraphy by mimicking some famous works of art) Plaintiff’s paintings. The court stated that “linmo” is fair use, and it is a common practice in the society. However, the user is required to (1) attribute work to the original author and indicate the name of the work; (2) the use must not infringe other’s exclusive rights.

- (2000) HuErZhongZhiChuZiDi 109 Hao civil decision:
  - Defendant’s website highly resembled Plaintiff’s website in the look and feel, structure, context, color, and font. The court stated that fair use is limited to personal use, and it excludes any commercial use. The scope must be limited, internal, and not published.
Are the Rules Applicable to DM or ML?

- Article 22: Enumerated Fair Use
  - (6) translation or reproduction in a small quantity of copies, of a published work for use by teachers or scientific researchers, in classroom teaching or scientific research, provided that the translation or reproduction shall not be published or distributed;
  - Most cases concern unlicensed copy and distribution of teaching materials
  - Uncertain for research purposes
    - Even if it applies, the research outcome or result cannot be published or distributed
Limitations on Economic Rights (Taiwan)

• Article 44 ~ Article 63 illustrations of limitations
  o Article 44 & 45:
    ▪ central or local government agencies may reproduce for the purpose of legislation, administration, and for judicial proceedings
  o Article 46 & 47:
    ▪ for the educational purposes, including teaching in schools, preparing teaching materials
  o Article 48 & 48-1
    ▪ for archival purposes of libraries, museums, and other cultural institutions
  o Article 49
    ▪ for News report purposes
Limitations on Economic Rights (Taiwan)

• Article 44 ~ Article 63 illustrations of limitations (cont.)
  o Article 52:
    ▪ Within a reasonable scope, works that have been publicly released may be quoted where necessary for reports, comment, teaching, research, or other legitimate purposes.
  o Article 59:
    ▪ The owner of a legal copy of a computer program may alter the program where necessary for utilization on a machine used by such owner, or may reproduce the program as necessary for backup; provided, this is limited to the owner’s personal use. If the owner referred to in the preceding paragraph loses ownership of the original copy for any reason other than the destruction or loss of the copy, all altered and backup copies shall be destroyed unless the economic rights holder grants its consent otherwise.
Limitations on Economic Rights (Taiwan)

• Article 65
  o Dual function
    ▪ Standard for interpreting the illustrations of limitations
      – “within a reasonable scope”
    ▪ Can be directly assert as a defense of an infringing act
Limitations on Economic Rights (Taiwan)

• Article 65
  o Fair use of a work shall not constitute infringement on economic rights in the work. In determining whether the exploitation of a work complies with the reasonable scope referred to in the provisions of Articles 44 through 63, or other conditions of fair use, all circumstances shall be taken into account, and in particular the following facts shall be noted as the basis for determination:
    1. The purposes and nature of the exploitation, including whether such exploitation is of a commercial nature or is for nonprofit educational purposes.
    2. The nature of the work.
    3. The amount and substantiality of the portion exploited in relation to the work as a whole.
    4. Effect of the exploitation on the work’s current and potential market value.
Where the copyright owner organization and the exploiter organization have formed an agreement on the scope of the fair use of a work, it may be taken as reference in the determination referred to in the preceding paragraph.
In the course of forming an agreement referred to in the preceding paragraph, advice may be sought from the specialized agency in charge of copyright matters.
Are the Rules Applicable to DM or ML?

• Probably in similar situation to the US
• Could (or should) the illustrations or enumeration of limitations become the more feasible legislative solution?
### Fair Use Factors / Illustrations / Enumerations

<table>
<thead>
<tr>
<th>Fair Use Factors</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>More flexible</td>
<td>Flexibility might create uncertainty, which is risky for AI researchers at schools, startups, and corporations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Illustrations / Enumerations</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>For lay person, it seems to be more predictable and trustworthy</td>
<td>Rigid rules may create interpretation and application problems for courts</td>
</tr>
</tbody>
</table>
Questions? Advice?

Thank you for your attention!

Gi-Kuen Jacob Li 李紀寬
Assistant Professor of Law
National Tsing Hua University, Taiwan
Institute of Law for Science and Technology
gkli@mx.nthu.edu.tw