Pairwise Learning to Rank Approcah with Neural Networks

*Note: Sub-titles are not captured in Xplore and should not be used

Andri Reimondo Tamba
Fakutlas Informatika dan Teknik Elektro
Institut Teknologi Del
Medan, Indonesia
andriraymondo@gmail.com

Martin Sinaga
Fakutlas Informatika dan Teknik Elektro
Institut Teknologi Del
Medan, Indonesia
enjelinhutahaean250@gmail.com

Enjelin Ida Hutahaean
Fakutlas Informatika dan Teknik Elektro
Institut Teknologi Del
Medan, Indonesia
line 5: email address or ORCID

Feber Manuel Sianipar
Fakutlas Informatika dan Teknik Elektro
Institut Teknologi Del
Medan, Indonesia
sianipar.feber@gmail.com

***Abstract*—**Pada penelitian ini kami menyajikan pendekatan pembelajaran berpasangan untuk menentukan peringkat berdasarkan neural net, yang disebut DirectRanker, yang menggeneralisasi RankNet artchitecture. Kami menunjukkan secara matematis bahwa model kami reflexive, antisymmetric, dan transitive yang memungkinkan untuk mensimplikasi training dan meningkatkan performa.

***Keywords*—** Informtaion Retrieval, Machine learning, Learning to rank.

I. INTRODUCTION

Information Retrieval atau temu kembali informasi merupakan salah satu aplikasi terpenting dari *machine learning techniques* beberapa tahun terakhir. Information retrieval merupakan sistem pencarian untuk menemukan kembali sebuah informasi. Tujuan dari *Information Retrieval* yaitu untuk mengambil informasi atau data serta menampilkan dokumen yang relevan. Jumlah data yang sangat besar yang digunakan pada kehidupan sehari-hari, seperti data penelitian, ekonomi, dan lain-lain. Sehingga hanya perlu untuk mengambil data yang relevan dengan kebutuhan user. Salah satu masalah utama dalam pencarian informasi adalah masalah *learning to rank*. Diberikan kueri q dan satu set dokumen d_1 ,…., d_n seseorang ingin menemukan *ranking* yang memberikan urutan dokumen sesuai dengan relevansinya tergantung pada *q (query)*.

Information retrieval akan mempermudah dalam pencarian data yang diutuhkan. Metode *machine leaning* merupakan metode yang paling umum dalam menangani pembelajaran untuk menentukan rank problem. Pada makalah ini, peneliti menggunakan file *artifical neural net*, dalam sepasang dokumen, untuk menemukan dokumen yang lebih relevan. Pendekatan ini dikenal sebagai pairwise ranking approach, yang kemudian dapat digunakan untuk mengurutkan daftar dokumen. *Architecture neural net* yang dipilih akan menampilkan sifat-sifat tertentu yang secara signifikan meningkatkan kinerja dibandingkan dengan pendekatan lain. Pada dasarnya struktur neural net sama dengan yang ada di RankNet. Namun beberapa kendala yang digunakan disana dan menggunakan algortima pengoptimalan yang lebih modern. Penelitian ini mengarah pada kinerja yang ditingkatkan secara signitifikan dan menempatkan penekatan kami head-to head ddengan metode yang lebih canggih. Hal ini sangat luar biasa mengigat struktur model yang relatif sederhana dan waktu pelatihan serta pengujian yang singkat. Lebih lanjut, peneliti menggunakan formasi yang berbeda untuk mendeskripsikan properti model kami dan menemukan bahwa model tersebut secara inheren, refleksif, entisimetris, dan transitif. Singkatnya, konribusi dari makalah ini adalah:

1. Kami mengusulkan skema sederhana dan efektif untuk *neural netwoek structures* untuk pairwise ranking, yang disebut DirectRanker yang merupakan generalisasi dari RankNet.
2. Analisis teoritis menunjukkan komponen mana dari struktur jaringan tersebut yang memunculkan propertinya dan apa persyaratan pada data pelatihan untuk membuatnya berfungsi.
3. Menjaga esensi RankNet dan mengoptimalkannya dengan metode modern, eksperimen menunjukkan bahwa, bertentangan dengan kepercayaan umum, metode berpasangan masih dapat bersaing dengan metode listwise yang lebih baru dan jauh lebih kompleks.
4. Menunjukkan bahwa metode tersebut dapat digunakan pada prinsipnya untuk klasifikasi. Secara khusus, untuk penemuan substruktur dalam satu kelas.

II. METODOLOGI

A. Pendekatan Pairwise

Pairwise ranking digunakan dengan cara membuat tabel dan membuat komparasi antara satu ide dengan yang lainnya berdasarkan variabel penentu keamanan pelaksanaan Dalam pendekatan *pairwise*, pasangan dokumen dianggap sebagai *input* untuk sistem pembelajaran. Tujuannya adalah bukan untuk menentukan skor relevansi setiap dokumen dalam kaitannya dengan kueri, tetapi dokumen mana yang lebih relevan daripada yang lain.

B. Neural Network

Neural network adalah arsitektur berbasis jaringan saraf, yang pada dasarnya terdiri dari beberapa tahapan, jaringan tersebut dapat dilatih untuk melakukan tugas-tugas terkait

klasifikasi teks. Berikut merupakan tahapan dari neural network adalah sebagai berikut (Aich, 2019):

1. Convolutional layers
   Lapisan konvolusional ini merupakan salah satu lapisan penting dalam CNN. Lapisan ini berisi sejumlah matriks kernel. Pada lapisan ini, konvolusional biasanya dilakukan oleh matriks kernel pada input dan keluaran sebagai matriks fitur nilai tambah bias dihasilkan. Bobot kernel dan bias dipelajari dengan menggunakan prosedur pembelajaran karena bobot koneksi dibagi di antara neuron.
2. Pooling layers
   Pooling layers merupakan elemen fundamental dari CNN. ujuan utama dari lapisan ini adalah untuk melakukan pengurangan dimensi input, yang mengurangi jumlah yang dihasilkan secara acak variabel sehingga proses analisis data lebih cepat dan sederhana. Subsampling dari lapisan konvolusi keluaran dilakukan oleh lapisan penyatuan dengan menggabungkan elemen tetangga. Fungsi penggabungan maksimal adalah fungsi penggabungan yang paling umum digunakan yang biasanya mengambil nilai maksimum di antara lingkungan lokal.
3. Embedding Layer
   Embedding layer ini merupakan lapisan khusus dari CNN yang berfungsi untuk melakukan tugas-tugas terkait klasifikasi teks. Tujuan dari lapisan ini adalah untuk mengubah dokumen teks masukan menjadi format yang tepat dan sesuai untuk CNN. Pada lapisan ini, setiap kata dari dokumen teks masukan diubah menjadi vektor padat ukuran tetap
4. Fully Connected Layer
   Fully connected layer merupakan lapisan tersembunyi dari FNN (*Feed-Forward*) dan juga merupakan lapisan terakhir yang paling banyak digunakan pada tahap CNN. Lapisan ini juga bisa sebagai lapisan konvolusi unik yang berisi matriks kernel dengan ukuran 1x1. Jenis lapisan ini adalah anggota grup yang berisi bobot lapisan yang bisa dilatih.
   Numerical Weight (Bobot Numerik) pada setiap masukan ditetapkan berdasarkan data historis sebelum proses pelatihan (*training process*). Dalam proses pelatihan, bobot optimal diselesaikan dengan mengurangi *mean square error* (E_m). Berikut merupakan rumus mencari *mean square error* (E_m):

$$E_m = \frac{1}{\mathrm{p}_1} * n_0 \sum_{p_1}^{p=p_1} * \sum_{s_1}^{s=n_0} e_s^2(p)$$

Keterangan :

Dimana, n_0 merupakan jumlah neuron dari lapisan keluaran dan e_s^2 merupakan error dari keluaran s_{th} untuk pola p_{th}set pelatihan. Untuk meminimalkan fungsi kesalahan e_m digunakan algoritma *mini-batch stochastic gradient descent* (m-SGD). Pada m-SGD ini koefisien model dan estimasi kesalahan model dilakukan dengan membagi *training set* menjadi sejumlah *batch* kecil. Algoritma m-SGD merupakan algoritma yang paling banyak digunakan di bidang *deep learning*.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

Identify applicable funding agency here. If none, delete this text box.

- Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and

variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a \square\square\square b \square\square\square\square\square \qquad \square\square\square$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Authors and Affiliations

The template is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

1) For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

2) For papers with less than six authors: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) Deletion: Delete the author and affiliation lines for the extra authors.

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1", "Heading 2", "Heading 3", and "Heading 4" are prescribed.

C. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. Table Type Styles

Table Head	Table Column Head		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a] Sample of a Table footnote. (*Table footnote*)

Fig. 1. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT *(Heading 5)*

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a