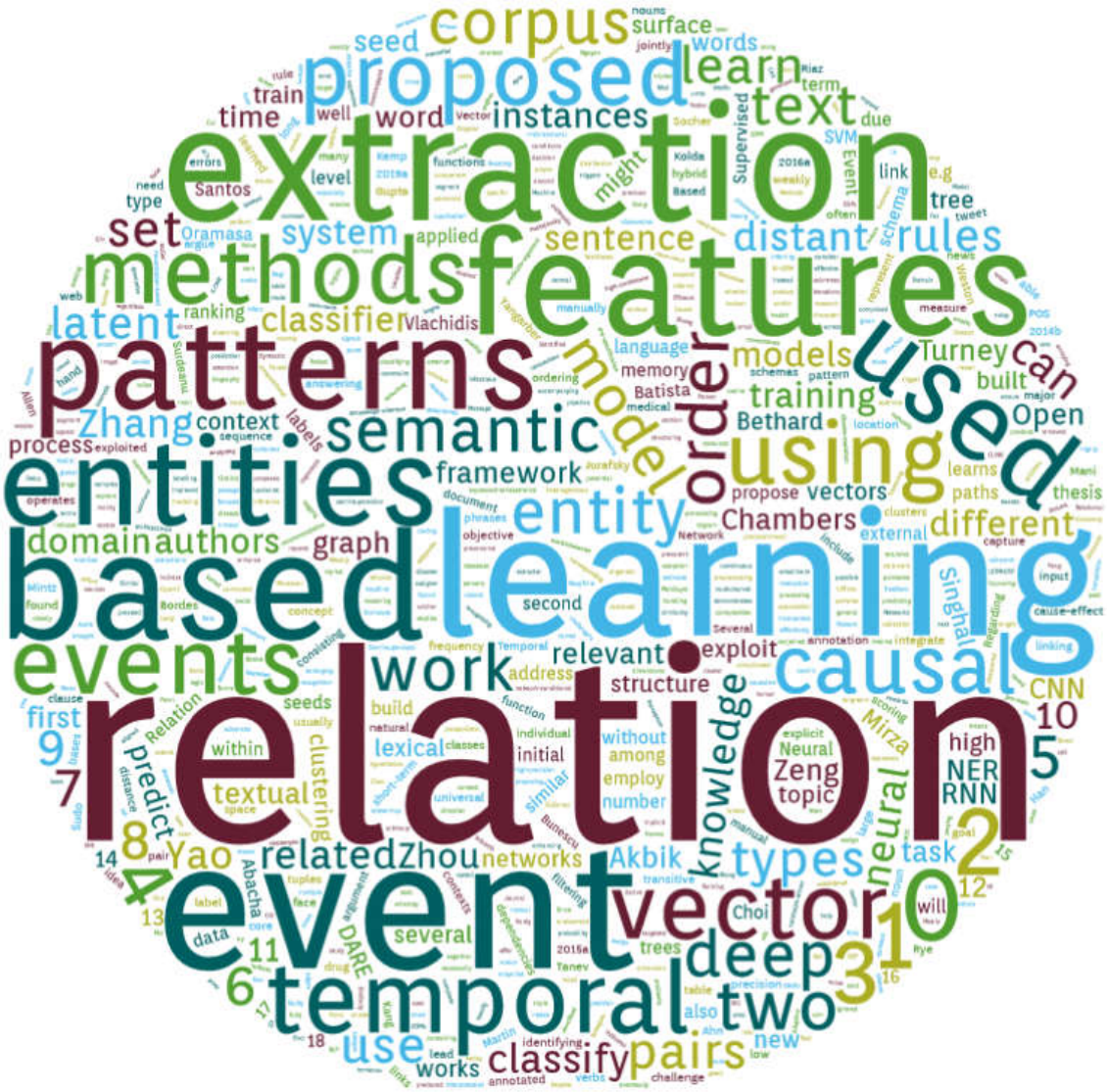
BASF
We create chemistry

**Technical Challenge**
**Patent Calssification**



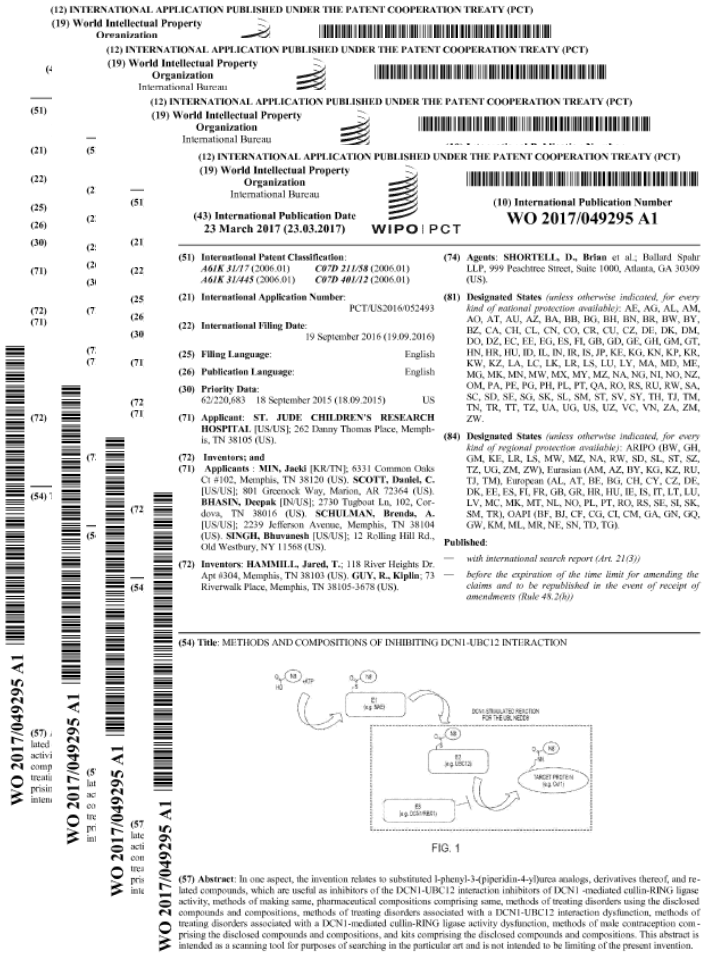


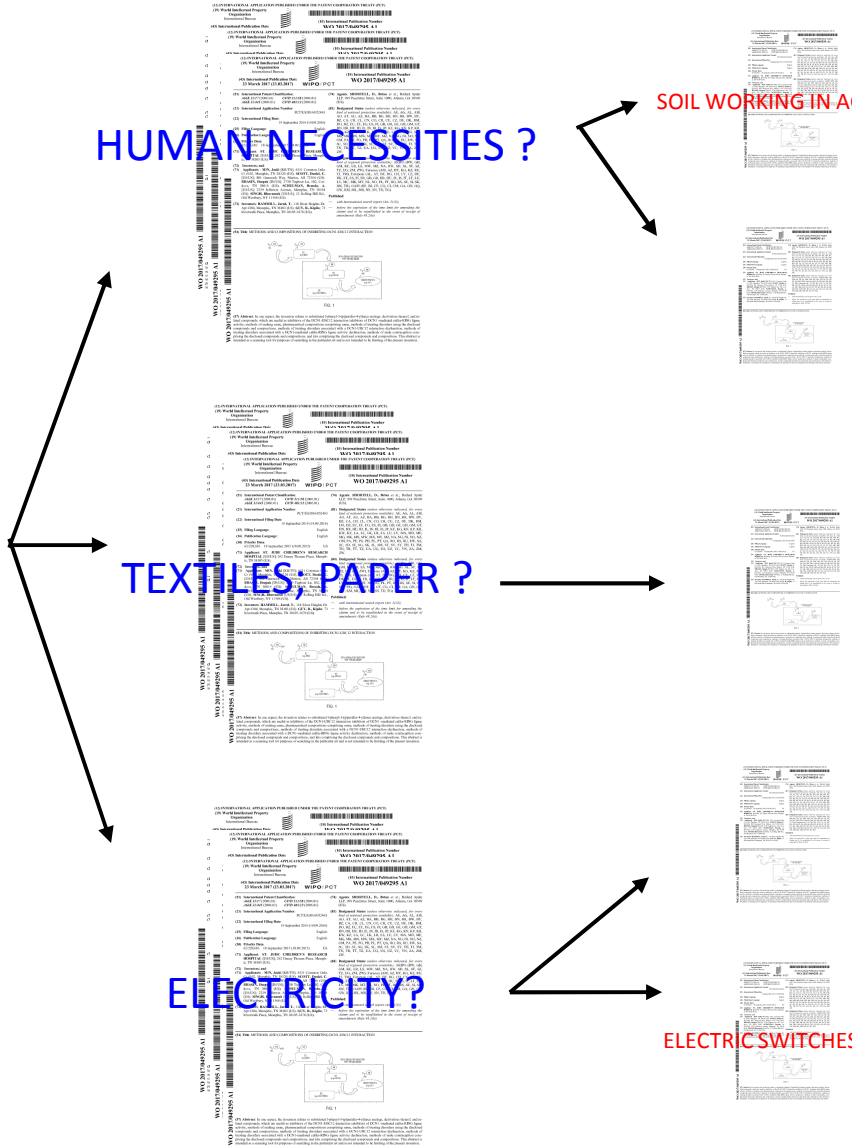
Duc-Thuan Vo, Ph. D
Postdoctoral Fellow

ls3 RYERSON
Laboratory for Systems, Software and Semantics

# Introduction

- Patent classification


(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)
(19) World Intellectual Property Organization
International Bureau
(10) International Publication Number
WO 2017/049295 A1
(43) International Publication Date
23 March 2017 (23.03.2017)
WIPO | PCT
(54) Title: METHODS AND COMPOSITIONS OF INHIBITING DCN1-UBC12 INTERACTION
FIG. 1
(57) Abstract: In one aspect, the invention relates to substituted 1-phenyl-3-(piperidin-4-yl)urea analogs, derivatives thereof, and related compounds, which are useful as inhibitors of the DCN1-UBC12 interaction inhibitors of DCN1-mediated cullin-RING ligase activity, methods of making same, pharmaceutical compositions comprising same, methods of treating disorders using the disclosed compounds and compositions, methods of treating disorders associated with a DCN1-UBC12 interaction dysfunction, methods of treating disorders associated with a DCN1-mediated cullin-RING ligase activity dysfunction, methods of male contraception comprising the disclosed compounds and compositions, and kits comprising the disclosed compounds and compositions. This abstract is intended as a scanning tool for purposes of searching in the particular art and is not intended to be limiting of the present invention.




LS3 RYERSON
Laboratory for Systems, Software and Semantics

# Methodology

- US patent structure

(12) **United States Patent**
**Bagheri et al.**

(10) **Patent No.: US 10,834,118 B2**
(45) **Date of Patent: Nov. 10, 2020**

(54) **AMBIGUITY RESOLUTION SYSTEM AND METHOD FOR SECURITY INFORMATION RETRIEVAL**

(71) Applicant: **International Business Machines Corporation**, Armonk, NY (US)

(72) Inventors: **Ebrahim Bagheri**, Toronto (CA); **Mohammadreza Barouni Ebrahimi**, Ottawa (CA); **Samaneh Bayat**, Ottawa (CA); **Zeinab Noorian**, Toronto (CA)

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

( * ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 227 days.

(21) Appl. No.: **15/837,061**

(22) Filed: **Dec. 11, 2017**

(65) **Prior Publication Data**

US 2019/0182285 A1 Jun. 13, 2019

(51) **Int. Cl.**
***H04L 29/06*** (2006.01)
***G06F 9/44*** (2018.01)
(Continued)

(52) **U.S. Cl.**
CPC ..... ***H04L 63/1433*** (2013.01); ***G06F 9/44*** (2013.01); ***G06F 11/362*** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC ... H04L 63/1433; H04L 63/1483; G06F 9/44; G06F 11/362; G06F 11/3672;

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,909,648 B2 * 12/2014 El-Yaniv ..... G06N 5/02 707/739
9,607,087 B1 * 3/2017 Ofek ..... G06F 16/951
(Continued)

OTHER PUBLICATIONS

Naresh Kumar Nagwani, Weight Similarity Measurement Model Based, Objected Oriented Approach for Bug Databases Mining to Detect Similar and Duplicate Bugs, dated Jan. 23-24, 2009, ACM, New York, pp. 202-207, 6 pages, (Year: 2017).*
(Continued)

*Primary Examiner* — Phong H Nguyen
(74) *Attorney, Agent, or Firm* — Yee & Associates, P.C.

(57) **ABSTRACT**

A method for ambiguity resolution in retrieving security information. A unified representation model utilizing the security information mined from a group of sources is generated by a computer system, wherein the unified representation model represents terms in the security information using vectors that describe locations in space for the terms. Distances between the terms indicate correlations between the terms, and the terms are for security concepts in the security information. The computer system represents the security concepts utilizing a group of weighted graphs that identifies semantic-relatedness between the terms utilizing the vectors in the unified representation model, wherein the group of weighted graphs enables interpreting queries with reduced ambiguity. The computer system interprets a seed query utilizing the group of weighted graphs. The computer system retrieves a portion of the security information utilizing the seed query interpreted utilizing the group of weighted graphs.

accordance with an illustrative embodiment.

DETAILED DESCRIPTION

The present invention may be a system, a method, and/or a computer program product. The computer program product may include a computer-readable storage medium (or media) having computer-readable program instructions thereon for causing a processor to carry out aspects of the present invention.

The computer-readable storage medium can be a tangible device that can retain and store instructions for use by an instruction execution device. The computer-readable storage medium may be, for example, but is not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. A non-exhaustive list of more specific examples of the computer-readable storage medium includes the following: a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a static random access memory (SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disk (DVD), a memory stick, a floppy disk, a mechanically encoded device such as punch-cards or raised structures in a groove having instructions recorded thereon, and any suitable combination of the foregoing. A computer-readable storage medium, as used herein, is not to be construed as being transitory signals per se, such as radio waves or other freely propagating electromagnetic waves, electromagnetic waves propagating through a waveguide or other transmission media (e.g., light pulses passing through a fiber-optic cable), or electrical signals transmitted through a wire.

Computer-readable program instructions described herein can be downloaded to respective computing/processing devices from a computer-readable storage medium or to an external computer or external storage device via a network, for example, the Internet, a local area network, a wide area network and/or a wireless network. The network may comprise copper transmission cables, optical transmission fibers, wireless transmission, routers, firewalls, switches, gateway computers and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer-readable program instructions from the network and forwards the computer-readable program instructions for storage in a computer-readable storage medium within the respective computing/processing device.

Computer-readable program instructions for carrying out operations of the present invention may be assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, or either source code or object code written in any combination of one or more programming languages, including an object oriented programming language such as Smalltalk, C++ or

area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider). In some embodiments, electronic circuitry including, for example, programmable logic circuitry, field-programmable gate arrays (FPGA), or programmable logic arrays (PLA) may execute the computer-readable program instructions by utilizing state information of the computer-readable program instructions to personalize the electronic circuitry, in order to perform aspects of the present invention.

Aspects of the present invention are described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer-readable program instructions.

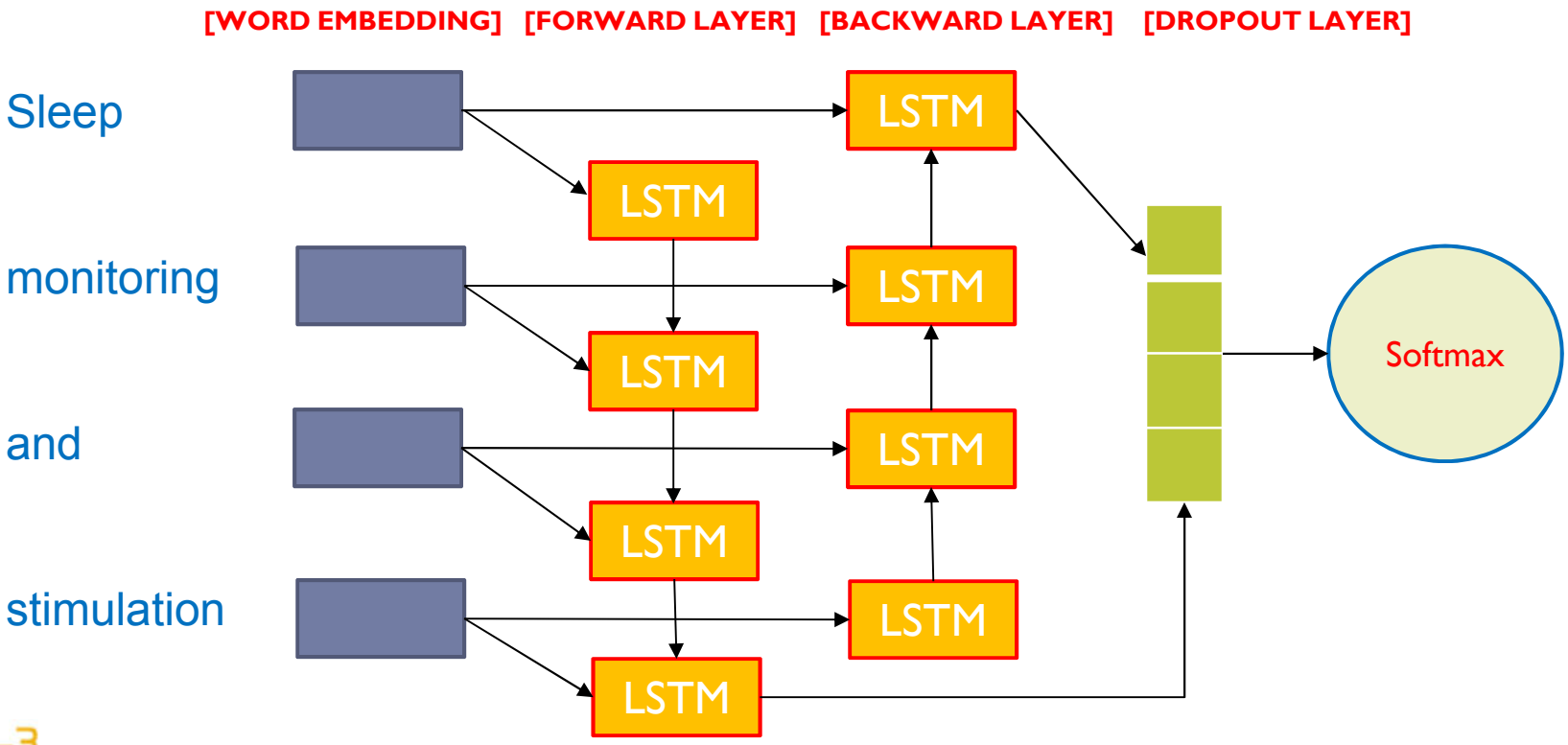
These computer program instructions may be provided to a processor of a general-purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer program instructions may also be stored in a computer-readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer-readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

The computer-readable program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other device to cause a series of operational steps to be performed on the computer, other programmable apparatus or other device to produce a computer implemented process, such that the instructions which execute on the computer, other programmable apparatus, or other device implement the functions/acts specified in the flowchart and/or block diagram block or blocks.

The flowchart and block diagrams in the figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods, and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, a segment, or a portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). In some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be

# Methodology

- Bidirectional Long Short Term Memory (RNN) (Liu et al., 2016)
  - Using sequential information
  - LSTM uses passing input from last ouput to retain information that able to leverage all information at the end to make predictions.
- Proposed model






Liu et al. (2016). Recurrent Neural Network for Text Classification with Multi-Task Learning. In IJCAI 2016

# Experimentation

- Experimental dataset
  - https://bulkdata.uspto.gov/data/patent/grant/redbook/fulltext/2020/ipg200107.zip
  - https://bulkdata.uspto.gov/data/patent/grant/redbook/fulltext/2020/ipg200114.zip
  - 12,344 patent records
- Categories

Main categories (8)

| | |
|---|---|
| **A** | HUMAN NECESSITIES |
| **B** | PERFORMING OPERATIONS; TRANSPORTING |
| **C** | CHEMISTRY; METALLURGY |
| **D** | TEXTILES; PAPER |
| **E** | FIXED CONSTRUCTIONS |
| **F** | MECHANICAL ENGINEERING; LIGHTING; HEATING; WEAPONS; BLASTING |
| **G** | PHYSICS |
| **H** | ELECTRICITY |

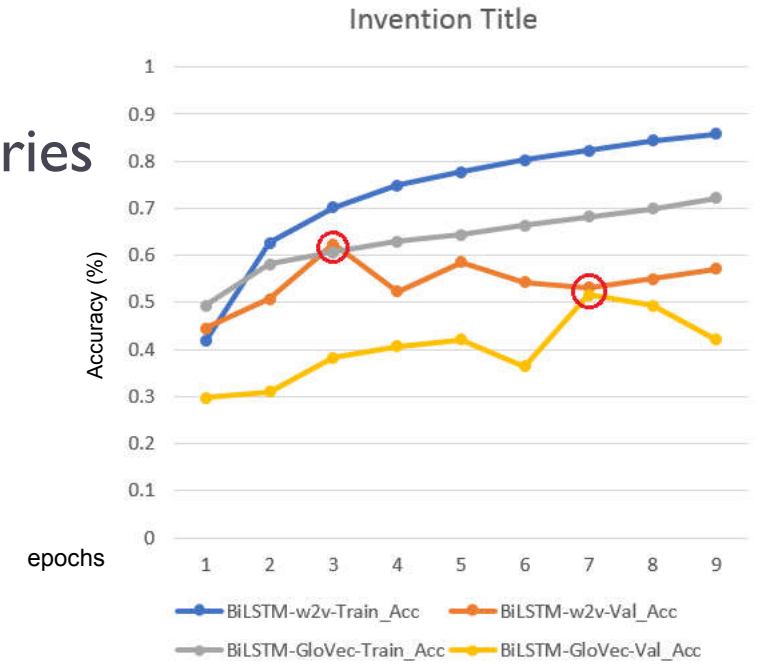
Sub-categories (478)

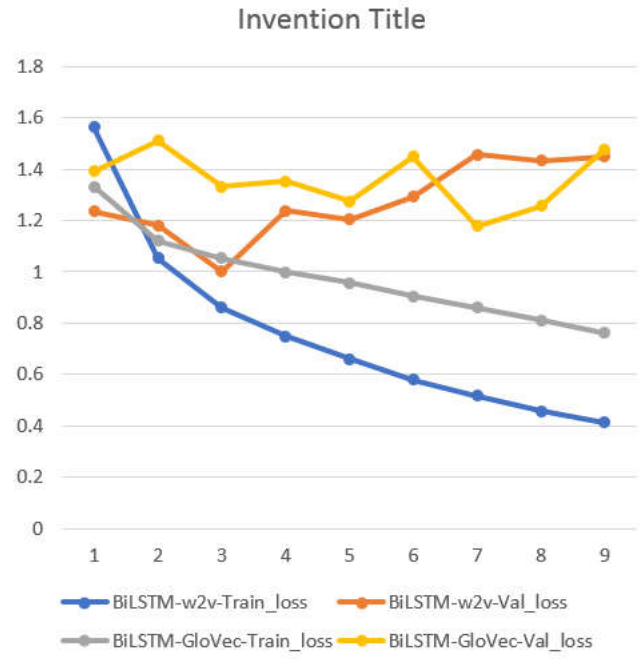
| | |
|---|---|
| **A01B** | SOIL WORKING IN AGRICULTURE OR FORESTRY … |
| **A01C** | PLANTING; SOWING; FERTILISING,… |
| **A01D** | HARVESTING; MOWING,… |
| **A01F** | PROCESSING OF HARVESTED PRODUCE … |
| **A01G** | HORTICULTURE,… |
| **A01H** | NEW PLANTS,... |
| **A01J** | MANUFACTURE OF DAIRY PRODUCTS,… |
| **A01K** | ANIMAL HUSBANDRY, ... |
| **...** | |
| **F01D** | MECHANICAL METHODS OR APPARATUS... |
| **...** | |
| **H01H** | ELECTRIC SWITCHES,... |

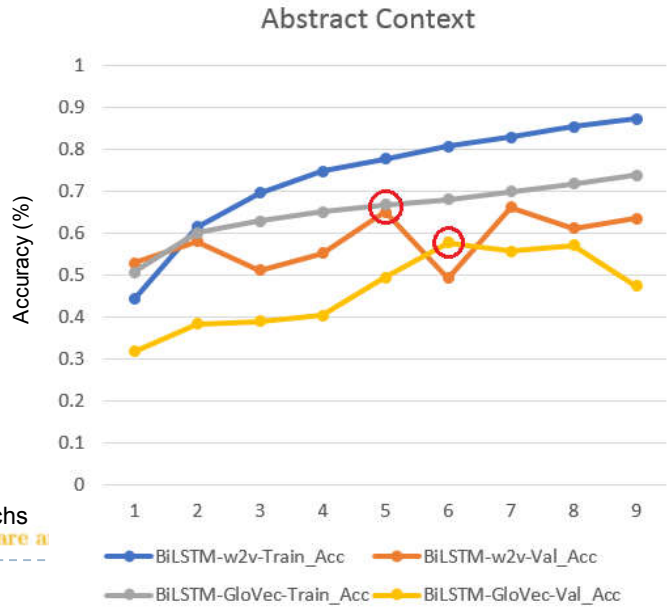
- Word Embedding
  - Word2Vec
  - GloVec

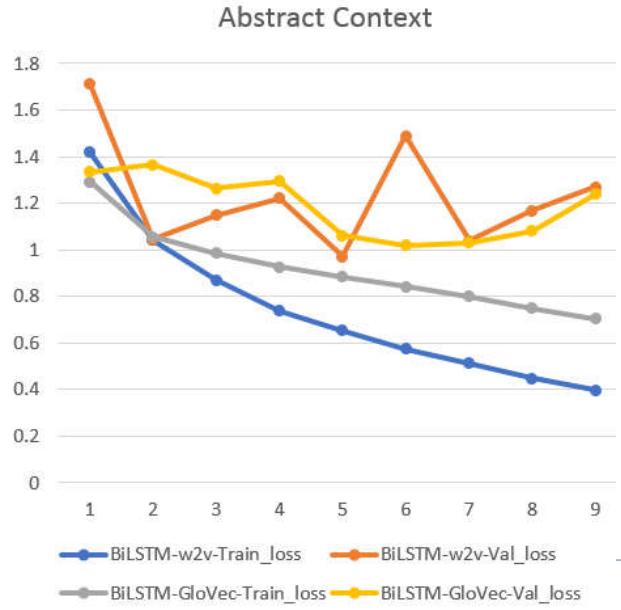
# Experimental results (Analysis)

- Main categories



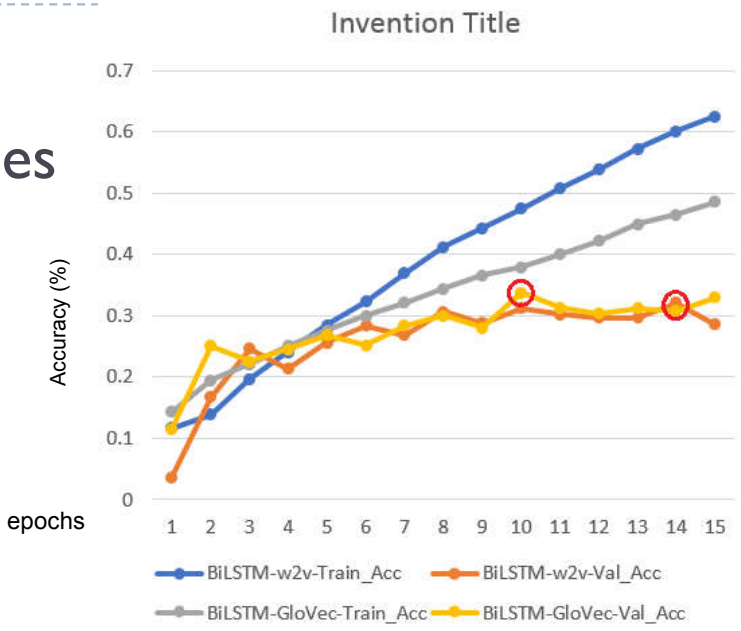



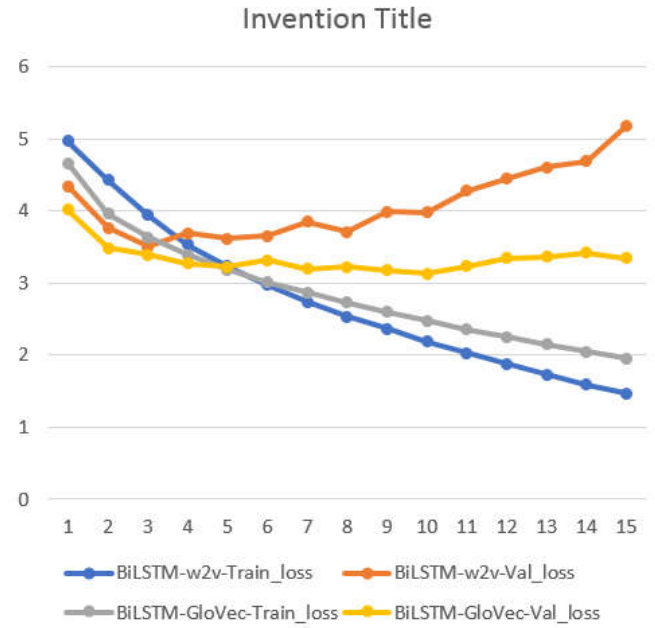



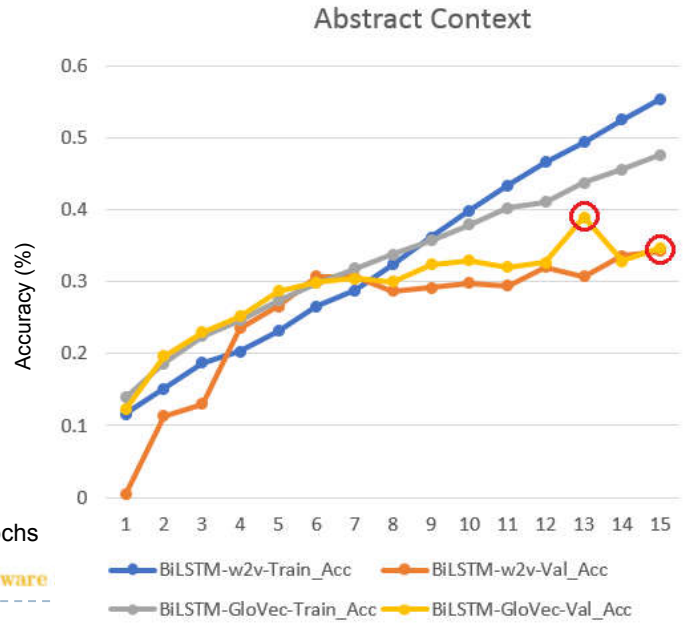

LS3 RYERSON Laboratory for Systems, Software a

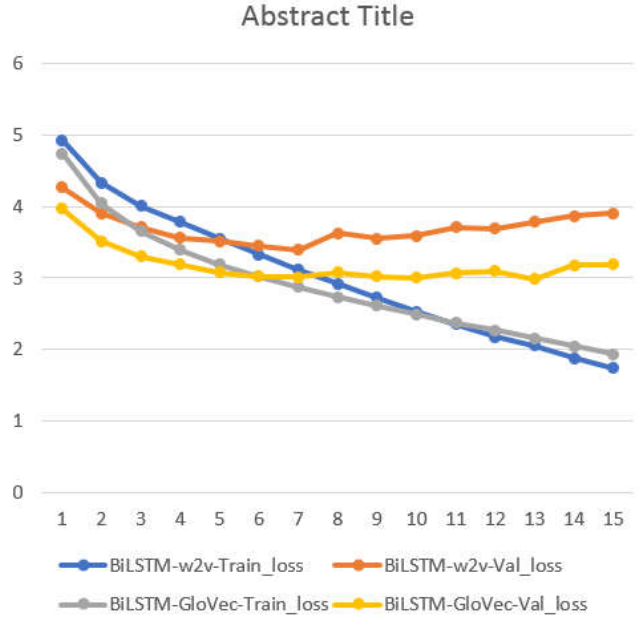
# Experimental results

- Sub categories




RYERSON Laboratory for Systems, Software

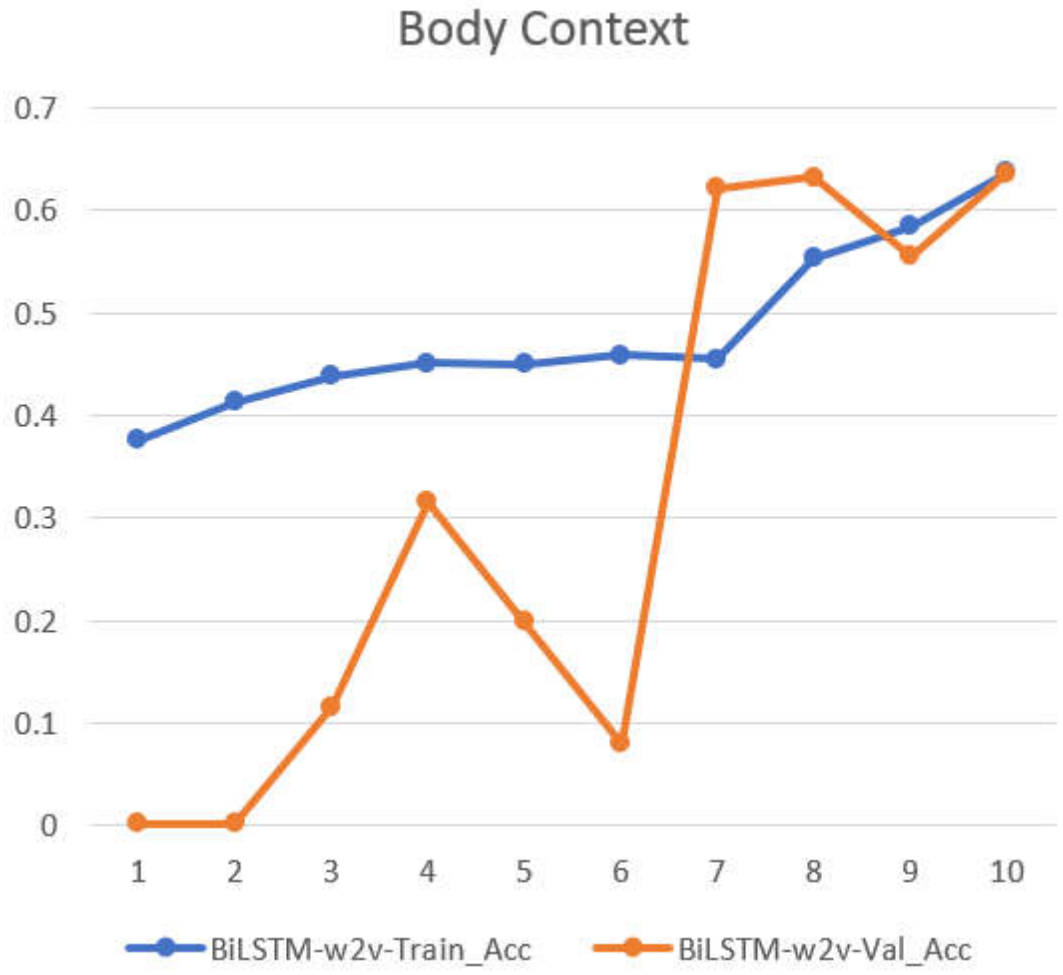
# Experimental results (15 Epochs)

- Main categories and sub categories (15 Epochs)

| | Main category (%) | Sub category (%) |
|---|---|---|
| **Invention Title** | | |
| BiLSTM-Word2Vec (Avg) | 54.43 | 25.88 |
| BiLSTM-Word2Vec (Max) | **62.37** | 32.13 |
| BiLSTM-GloVec (Avg) | 42.31 | 27.51 |
| BiLSTM-GloVec (Max) | 52.31 | **33.80** |
| **Abstract Context** | | |
| BiLSTM-Word2Vec (Avg) | 58.54 | 25.59 |
| BiLSTM-Word2Vec (Max) | **66.20** | 34.23 |
| BiLSTM-GloVec (Avg) | 47.26 | 29.04 |
| BiLSTM-GloVec (Max) | 57.73 | **38.93** |

# Discussion

- Patent classification based on body context



# Discussion

- Patent classification based on body context
  - BiLSTM is good? unigram and bigram?
  - Traing time
  - Body context (summary) ?
  - Tf*Idf
- Compare with traditional ML classification
  - SVM,
  - kNN
  - …
- Patent Storage
  - Hadoop
  - Lucene
  - …

Thank you for your attention.
Q&A