Distillation of Knowledge from the Research Literatures on Alzheimer’s Dementia

Wutthipong Kongburan, Mark Chignell, and Jonathan H. Chan

School of Information Technology
King Mongkut's University of Technology Thonburi
INTRODUCTION

PROPOSED METHOD

EXPERIMENT, RESULT AND DISCUSSION

CONCLUSION AND FUTURE WORK
Agenda

INTRODUCTION

PROPOSED METHOD

EXPERIMENTAL RESULT AND DISCUSSION

CONCLUSION AND FUTURE WORK
Alzheimer Disease

It’s the only cause of death in the top 10 in America that CANNOT BE PREVENTED, CURED OR SLOWED

AMOST TWO THIRDS of Americans with Alzheimer’s disease are women.

1 IN 3 SENIORS dies with Alzheimer’s or another dementia.

6 Alzheimer’s disease is the 6TH LEADING CAUSE OF DEATH IN THE UNITED STATES.

Only 45% of people with ALZHEIMER’S disease or their caregivers report BEING TOLD OF THEIR DIAGNOSIS.

More than 90% of people with the four most common types of CANCER have been TOLD OF THEIR DIAGNOSIS.

By 2050, these costs could rise as high as $1.1 TRILLION.

In 2015, Alzheimer’s and other dementias will cost the nation $226 BILLION.
Number of publications related to Alzheimer treatment added into PubMed every year
Text Mining Research & Applications

“the process of automatic deriving high-quality information from natural language text”

What is Sentiment Analysis?
A linguistic analysis technique that identifies opinion early in a piece of text.

The movie is great. 😊
The movie stars Mr. X 😞
The movie is horrible. 😞
Overview of Information Extraction

Omega-3 fatty acids in fish may help prevent cognitive decline in Alzheimer patient.

Omega-3 fatty acids in fish may help prevent cognitive decline in Alzheimer patient.
Named Entity Recognition (NER)

- Classify tokens into predefined categories such as person name, organizations, diseases, proteins.

Stanford Named Entity Recognizer for treatment of mild to moderate Alzheimer disease, a randomized controlled trial.

Context: Several reports from small clinical trials have suggested that estrogen replacement therapy may be useful for the treatment of Alzheimer disease (AD) in women.

Objective: To determine whether estrogen replacement therapy affects global, cognitive, or functional decline in women with mild to moderate AD.


Setting: Thirty-two study sites in the United States.

Participants: A total of 120 women with mild to moderate AD and a Mini-Mental State Examination score between 12 and 28 who had had a hysterectomy. Interventions: Participants were randomized to estrogen, 0.625 mg/d (n = 42), or 1.25 mg/d (n = 39), or to identically appearing placebo (n = 39). One subject withdrew after randomization but before receiving medication; 97 subjects completed the trial.

Main Outcome Measures: The primary outcome measure was change on the Clinical Global Impression of Change (CGIC) 7-point scale, analyzed by intent to treat; secondary outcome measures included other global measures as well as measures of mood, specific cognitive domains (memory, attention, and language), motor function, and activities of daily living; compared by the combined estrogen groups vs the placebo group at 2, 6, 12, and 15 months of follow-up.

Results: The CGIC score for estrogen vs placebo was 5.1 vs 5.0 (P = .43); 80% of participants taking estrogen vs 74% of participants taking placebo worsened (P = .49). Secondary outcome measures also showed no
## NER Features and Models

<table>
<thead>
<tr>
<th>Features</th>
<th>Gene and Protein</th>
<th>Chemical</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(gene/protein and disorders)</td>
</tr>
<tr>
<td>Name</td>
<td>Gimli</td>
<td>NERSuite</td>
<td>IBM Watson</td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td></td>
<td>[37]</td>
</tr>
<tr>
<td>Programming Language</td>
<td>Java</td>
<td>C++</td>
<td>-</td>
</tr>
<tr>
<td>Open source</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

### Features

- **Linguistic**
  - Normalization
  - POS
  - Chunking
  - Dependency

- **Orthographic**
  - Capitalization
  - Counting
  - Symbols

- **Morphological**
  - Suffix and Prefix
  - Charng-grams
  - Word Shape

- **Context**
  - Windows
  - Conjunctions

- **Lexicons**
  - Target names
  - Trigger names

### Model

- Supervised
- Semi-supervised
- ASO

### Model Combination

<table>
<thead>
<tr>
<th>Post-processing</th>
<th>CRF</th>
<th>CRF</th>
<th>CRF</th>
<th>CRF</th>
<th>CRF</th>
<th>CRF</th>
<th>CRF</th>
<th>CRF</th>
<th>HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parentheses</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexicons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Our Features

- Word features: current word, previous word, next word, all words within a window
- Orthographic features:
  - Jenny → Xxxx
  - IL-2 → XX-#
- Prefixes and Suffixes:
  - Jenny → <J, <Je, <Jen, ..., nny>, ny>, y>
- Label sequences
- Lots of feature conjunctions
Conditional Random Fields (CRFs)

a type of discriminative probabilistic model

- O is observation words sequence \( o = \langle o_1, o_2, \ldots, o_n \rangle \)
- S is label sequence \( s = \langle s_1, s_2, \ldots, s_n \rangle \)

Linear chain CRFs define the conditional probability of a state sequence given an input sequence to be:

\[
P(s|o) = \frac{1}{Z_o} \exp \left( \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_j f_j(s_{i-1}, s_i, o, i) \right)
\]

- \( Z_o \) is a normalization factor of all state sequences [0, 1]
- \( f_j(s_{i-1}, s_i, o, i) \) is one of \( m \) functions that describes a feature
- \( \lambda_j \) is a learned weight for each such feature function (define)
  - If \( \lambda_j < 0 \) : CRFs to avoid using this feature
  - If \( \lambda_j = 0 \) : this feature does not affect the probability
  - If \( \lambda_j > 0 \) : improve probability of \( S \)
- \( s_{i-1}, s_i \) represent previous and current label for observation words position \( i \)
Conditional Random Fields (CRFs)

\[
P(s|o) = \frac{1}{Z_0} \exp \left( \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda f_j(s_{i-1}, s_i, o, i) \right)
\]

- \( f_j(s_{i-1}, s_i, o, i) \)
  - \( 1 \) if \( s_i \) = Intervention and \( o \) begins with memantin-
  - \( 0 \) Otherwise
- \( \lambda = 1 \)
- \( Z_0 \) is a normalization factor

Morphological feature
Prefix
memantin-....

Extract feature/
Train model

Testing token
memantinchlorid

Classifier/
Model

NER

P(Disease | memantinchlorid) = 0.0
P(Other | memantinchlorid) = 0.0
P(Intervention | memantinchlorid) = 1.0

memantinchlorid Intervention

Training Corpus

memantine
memantine monotherapy
memantine treatment

Intervention

Intervention

Intervention
Distillation of Knowledge from the Research Literatures on Alzheimer’s Dementia

“Text mining technology may assist in distilling knowledge from the vast corpus of research literature on Alzheimer’s dementia.”
Agenda

INTRODUCTION

PROPOSED METHOD

EXPERIMENTAL RESULT AND DISCUSSION

CONCLUSION AND FUTURE WORK
Donepezil can be prescribed for people with early to mid-stage Alzheimer's disease.

- AD Intervention corpus
  - Manual annotation
  - Medical abstracts
  - Train classifier
  - Named entity recognition
  - Annotation results
  - Comparison
  - Evaluation results
  - Manually annotated sentence

```
<table>
<thead>
<tr>
<th>donepezil</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>can</td>
<td>0</td>
</tr>
<tr>
<td>be</td>
<td>0</td>
</tr>
<tr>
<td>prescribed</td>
<td>0</td>
</tr>
<tr>
<td>for</td>
<td>0</td>
</tr>
<tr>
<td>people</td>
<td>0</td>
</tr>
<tr>
<td>with</td>
<td>0</td>
</tr>
<tr>
<td>early</td>
<td>0</td>
</tr>
<tr>
<td>to</td>
<td>0</td>
</tr>
<tr>
<td>mid-stage</td>
<td>0</td>
</tr>
<tr>
<td>Alzheimer's Disease</td>
<td>0</td>
</tr>
<tr>
<td>disease</td>
<td>0</td>
</tr>
<tr>
<td>Disease</td>
<td>0</td>
</tr>
</tbody>
</table>
```
Corpus construction 50 abstracts

“Alzheimer treatment”
20 abstracts

“Alzheimer treatment”
10 abstracts

“Nonpharmacologic treatment for Alzheimer”
20 abstracts
The three labels assigned are Disease, Intervention, and Other.

All entities to be annotated were nouns.

A general term will be excluded.

Pronouns and co-references were excluded.

Special characters (e.g., quote, dash, or parenthesis) at the beginning or end of entities were not considered.
Disease included both synonymous mentions and abbreviated forms of AD.
Any object or action that is performed by doctor or other clinician targeted in order to help the AD patient.

- Treatment, prevention, diagnosis, cause, risk factor

- Such typical treatment - the side effects.
- Medical interventions
  - alternative treatment
  - increasing patient satisfaction
  - lead to better health outcomes
Donepezil can be prescribed for people with early to mid-stage Alzheimer's disease.

AD Intervention corpus → Train classifier → Named entity recognition → Annotation results → Comparison → Evaluation results

Manual annotation
Medical abstracts

Manually annotated sentence

donepezil Intervention
can 0
be 0
prescribed 0
for 0
people 0
with 0
early 0
to 0
mid-stage 0
Alzheimer's Disease 0
disease 0
Disease 0
“Alzheimer treatment”

Experiment (1)

first 10 webpages

AD Intervention corpus → Train classifier → Named entity recognition → Annotation results → Comparison → Evaluation results

Manual annotation → Medical abstracts → Train classifier

Manually annotated sentence

donepezil Intervention
can 0
be 0
prescribed 0
for 0
people 0
with 0
early 0
to 0
mid-stage 0
Alzheimer's Disease 0
disease 0
Disease . 0
<table>
<thead>
<tr>
<th>Measure</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.987</td>
</tr>
<tr>
<td>Recall</td>
<td>0.397</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.566</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.982</td>
</tr>
</tbody>
</table>

**Experiment (1)**

- Different antipsychotic drugs
- Cardiovascular disease environmental domains
- Memantine
- Behavior therapy
- Cognitive therapy
- Cholinesterase inhibitors include donepezil
- Music cognitive behavioural therapy
- Cueing taking memantine Chinese medicine
- Multi-sensory therapy Multi-sensory stimulation Ginkgo biloba
- Alzheimer’s drugs Caprylic acid environmental
- Olanzapine Regular exercise immunization therapy
- Coral calcium eicosapentaenoic acid cognitive stimulation

**Cholinesterase inhibitors**

- Animal-assisted therapy
- Vitamin
- Exercise
- Psychological therapy
- Antipsychotic drugs
- Dance therapy
- Huperzine A AChE inhibitors
- Massage
- Music therapy Kinesiotherapy
- Unfamiliar music Validation therapy
- Antipsychotics
- Strengthening routine habits
- Reminiscence therapy
Discussion

- Further analysis on confusion matrix, false positive and false negative tokens of intervention entity
  - errors between intervention and other class -> such intervention entities are not in corpus (e.g., Aromatherapy, relaxing song, art therapy, coconut oil, pet therapy, Rivastigmine or Exelon, Galantamine or Razadyne).
  - errors between intervention and disease class -> entity's boundary. For instance, Disease Alzheimer’s should be labeled as Disease, however when the token is Disease Alzheimer’s drugs, they must be labeled as Intervention.
  - a case sensitive in configuration. Although we can exclude a proper noun from annotation, such as Alzheimer’s Healthcare Center, a mislabeled entity can occur for example, Memantine is not annotated as Intervention.
### Experiment (2)

<table>
<thead>
<tr>
<th>Search criteria</th>
<th>Unseen/new Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal search</td>
<td>unfamiliar music, dance therapy, environmental cueing, Chinese medicine, coral calcium, animal-assisted therapy, multi-sensory therapy, massage</td>
</tr>
<tr>
<td>Past week</td>
<td>physical therapy, intranasal insulin therapy</td>
</tr>
<tr>
<td>Past month</td>
<td>physical therapy, intranasal insulin therapy, antiepileptic drugs, light therapy, flashing light therapy, mouse’s spine</td>
</tr>
<tr>
<td>Past year</td>
<td>flashing light therapy, routine screening biogen therapy, nilvadipine, BACE inhibitors</td>
</tr>
</tbody>
</table>
INTRODUCTION

PROPOSED METHOD

EXPERIMENTAL RESULT AND DISCUSSION

CONCLUSION AND FUTURE WORK
Conclusion and Future work
THANK YOU FOR YOUR ATTENTION ANY QUESTIONS

Publications

Relationships

Structured Information