University of Qom

Engineering Management and Soft Computing

JEMSC.QOM
Online ISSN: 2538-2675

An Algorithm for Fuzzification of WordNets and its Application

Yousef Alizadeh-Q[1], Behrouz Minaei-Bidgoli[2]✉, Sayyed-Ali Hossayni[3], Mohammad-R Akbarzadeh-T[4], Diego Reforgiato Recupero[5] and Aldo Gangemi [6]

1. Ph.D. Student, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran. Email: alizadeh.usef@gmail.com
2. Corresponding author, Prof., School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran. Email: b_minaei@iust.ac.ir
3. Postdoc Graduated, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran. Email: hossayni@iran.ir
4. Prof., Departments of Electrical Engineering and Computer Engineering, Ferdowsi University of Mashhad,Iran. Email: akbazadeh@ieee.org
5. Associate Prof., Department of Math. and Computer Science, University of Cagliari, Italy. Email: diego.reforgiato@unica.it
6. Prof., University of Bologna and ISTC, National Research Council, Italy. Email: aldo.gangemi@unibo.it

Article Info

Article type:
Research Article

Article history:
Received 2023 October 10
Received in revised form 2023 December 21
Accepted 2024 January 19
Published online 2024 March 15

Keywords:
Artificial intelligence,
Deep learning,
Electronic government,
Information technology,
Web service.

ABSTRACT

Electronic government refers to providing continuous government information and services to people through the Internet or other digital methods. The electronic government revolution can transform the public sector and re-establish the relationship between the government and citizens. This revolution was only practical with the element called artificial intelligence. One of the most significant developments of the last few years is the launch of e-government by artificial intelligence worldwide. In this paper, we address the challenges of e-government systems and propose a framework that uses artificial intelligence. In the next step, we present an intelligent e-government database architecture that supports developing and implementing e-government artificial intelligence applications. The main goal is to use reliable artificial intelligence techniques to advance the current e-government state. Services to minimize processing time, reduce costs, and improve citizen satisfaction.

Cite this article: Alizadeh-Q, Y., Minaei-Bidgoli, B. Hossayni, S.A. Akbarzadeh-T, M.R. Reforgiato Recupero, D. & Gangemi, A. (2023). An Algorithm for Fuzzification of WordNets and its Application. *Enginiiring Management and Soft Computing, 9 (2)*. 119-131.
DOI: https://doi.org/

© The Author(s)
DOI: https://doi.org/

Publisher: University of Qom

JEMSC.QOM
شاپا الکترونیکی: ۲۶۷۵-۲۵۳۸

مدیریت مهندسی و رایانش نرم

دانشگاه قم

ارائه الگوریتمی جهت فازی کردن وردنت و کاربرد آن در تحلیل احساسات

یوسف علیزاده غنی‌آبادی[۱]، بهروز مینایی بیدگلی[۲]✉، سید علی حسینی[۳]، محمدرضا اکبرزاده توتونچی[۴]، دیگو رفورجاتو رکوپرو[۵] و آلدو گانجمی[۶]

۱. دانشجوی دکتری، مهندسی کامپیوتر، دانشکده کامپیوتر دانشگاه علم و صنعت، تهران، ایران. رایانامه: alizadeh.usef@gmail.com

۲. نویسنده مسئول، استاد، مهندسی کامپیوتر، دانشکده کامپیوتر، دانشگاه علم و صنعت، تهران، ایران. رایانامه: b_minaei@iust.ac.ir

۳. پسادکتری، مهندسی کامپیوتر، دانشکده کامپیوتر دانشگاه علم و صنعت، تهران، ایران. رایانامه: hossayni@iran.ir

۴. استاد، مهندسی کامپیوتر، دانشکده کامپیوتر دانشگاه فردوسی، مشهد، ایران. رایانامه: akbazadeh@ieee.org

۵. دانشیار، علوم کامپیوتر، دانشکده ریاضیات و علوم کامپیوتر، دانشگاه Cagliari، ایتالیا. رایانامه: diego.reforgiato@unica.it

۶. استاد، علوم کامپیوتر، دانشکده فلسفه دانشگاه Council ، ایتالیا. رایانامه: aldo.gangemi@unibo.it

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۰۸/۰۱

تاریخ بازنگری: ۱۴۰۲/۰۹/۳۰

تاریخ پذیرش: ۱۴۰۲/۱۰/۲۹

تاریخ انتشار: ۱۴۰۲/۱۲/۲۵

کلیدواژه‌ها:
مدیریت عدم‌قطعیت،
تحلیل احساسات،
نظریه امکان،
وردنت فازی.

چکیده

پایگاه‌های داده واژگانی شبه‌وردنت (WLD) کلمات انگلیسی را در مجموعه‌ای از مترادف‌ها به نام هم‌نشیم «Synsets» گروه‌بندی می‌کنند. اگرچه WLDهای استاندارد در بسیاری از برنامه‌های کاربردی موفق متن‌کاوی مورداستفاده قرار می‌گیرند، اما این محدودیت را دارند که حس کلمه به‌عنوان نمایش‌دهنده معنای مرتبط با هم‌نشیمهای متناظر آنها به یک‌میزان در نظر گرفته می‌شود که به‌طورکلی درست نیست. به‌منظور غلبه بر این محدودیت، چندین نسخه فازی از هم‌نشیم ارائه شده است. یکی از ویژگی‌های مشترک این مطالعات این است که هدف آن‌ها تولید نسخه‌های فازی شده از WLDهای موجود نیست، بلکه ساخت WLDهای جدید را از ابتدا انجام می‌دهند. این کار در کاربردهایی که منابع و برنامه‌های کاربردی آن بر اساس WLDهای موجود است در متن‌کاوی محدودیت‌های زیادی را به وجود آورده است. در این مطالعه، ما الگوریتمی را برای ساخت نسخه‌های فازی WLD از هر زبان، باتوجه‌به مجموعه‌ای از اسناد و یک سیستم ابهام‌زدایی حس کلمه (WSD) برای آن زبان ارائه می‌کنیم. سپس، با استفاده از Open-American-National-Corpus و UKB WSD به‌عنوان ورودی‌های الگوریتم، نسخه فازی شده WordNet انگلیسی (FWN) را ساخته و به‌صورت آنلاین منتشر می‌کنیم. ما همچنین یک اثبات عملی برای اعتبار نتایج آن پیشنهاد می‌کنیم.

استناد: علیزاده غنی‌آبادی؛ مینایی بیدگلی، بهروز؛ حسینی، سیدعلی؛ اکبرزاده توتونچی، محمدرضا؛ رفورجاتو رکوپرو، دیگو و گانجمی، آلدو. (۱۴۰۲). «ارائه الگوریتمی جهت فازی کردن وردنت و کاربرد آن در تحلیل احساسات». *مدیریت مهندسی و رایانش نرم، دوره ۹ (۲)*. صص: ۱۳۱-۱۱۹.
https://doi.org/

ناشر: دانشگاه قم

© نویسندگان.

۱) مقدمه

با گسترش هر چه بیشتر استفاده از اینترنت و فضای مجازی، تولید محتوای متنی نیز بیشتر شده (چو و جیمز، ۲۰۲۱) و متعاقباً نیاز به انجام تحلیل روی این محتوا اهمیت بیشتری پیدا کرده است (حسینی و همکاران، ۲۰۲۰). پردازش زبان‌های طبیعی یا متن‌کاوی به‌عنوان یکی از شاخه‌های هوش مصنوعی تلقی می‌شود که هدف آن تحلیل خودکار متون و استخراج اطلاعات از آن است (هرشکوویچ و دوناتلی، ۲۰۲۱). لذا، در سال‌های اخیر توجه به این شاخه افزایش چشمگیری داشته است (خورانا و همکاران، ۲۰۲۳).

یکی از زیرشاخه‌های متن‌کاوی، تحلیل احساسات متون است (بیرجالی و همکاران ۲۰۲۱). تجزیه‌وتحلیل حجم عظیم اطلاعات بیان‌شده توسط افراد در شبکه‌های اجتماعی و سایت‌ها، برای شرکت‌ها، سیاستمداران و... بسیار مهم است (جاجی‌علی، ۲۰۲۰). تحلیل خودکار این متون اینترنتی که به‌وسیله کاربران تولیدشده است در بسیاری از کاربردها مانند تحلیل‌های اجتماعی، تجاری و سیاسی اهمیت بالایی دارد (شارما و جاین، ۲۰۲۰). اما ازآنجاکه کلمات را می‌توان به شکل‌ها و قالب‌های متعدد بیان کرد، استخراج اطلاعات مفید از این داده‌های بدون ساختار، یک مشکل جدی است (بندر و کولر، ۲۰۲۰) و پیداکردن یک رفتار مشخص و باقاعده برای آن مشکل خواهد بود (خورانا و همکاران، ۲۰۲۳).

ابتدا به بررسی نقش عدم‌قطعیت در متن‌کاوی و به طور خاص در تحلیل احساسات می‌پردازیم. پایگاه‌های داده واژگانی همچون وردنت[۷۳] که کلمات یک زبان را در گروه‌هایی هم‌معنی با هم –به نام هم‌نشیم[۷۴]– سازمان‌دهی می‌کنند، در روش‌های ماشینی، برای تحلیل احساسات استفاده می‌شود. در پایگاه‌داده‌های واژگانی رایج مانند وردنت، تمام اعضای هم‌نشیم‌ها به یک میزان عضو یک هم‌نشیم فرض می‌شوند؛ یعنی معنای آن هم‌نشیم را در یک سطح انتقال می‌دهند. به‌عبارت‌دیگر، این پایگاه‌داده‌های واژگانی فرض می‌کنند که هم‌نشیم‌ها مجموعه‌های قطعی و غیرفازی هستند. این فرض همیشه درست نیست و واژه‌بارهای مختلف از یک هم‌نشیم، ممکن است به میزان متفاوتی معنای هم‌نشیم خود را نشان دهند. این نقص روش‌های تحلیل احساسات مبتنی بر هم‌نشیم که از پایگاه‌داده‌های واژگانی با روش‌های تحلیل احساسات مبتنی بر واژگان استفاده می‌کنند، را نیز تحت‌تأثیر قرار می‌دهد، زیرا آن‌ها از همان هم‌نشیم‌های قطعی (غیرفازی) استفاده می‌کنند. روش‌های تحلیل احساسات مبتنی بر هم‌نشیم نیز وجود دارند که به همین ترتیب تحت‌تأثیر این مسئله قرار می‌گیرند

برای رفع مشکل مذکور، این پژوهش الگوریتمی را تدوین می‌کند که وردنت‌های موجود را بتواند فازی‌سازی نماید یا به‌عبارت‌دیگر به واژه‌بارهای عضو هم‌نشیم‌های یک وردنت موجود بتواند درجه عضویت اختصاص دهد که معادل خواهد بود با درجه سازگاری واژه‌بارها با هم‌نشیم‌های متناظر با آن واژه‌بار. بر اساس استدلالی که در بخش بعد ارائه خواهد شد، پیش‌بینی می‌شود در صورت ارائه الگوریتم پیشنهاد شده، صحت فرایندهای پردازش متن از جمله فرایند تحلیل احساسات افزایش یابد.

۲) پیشینه پژوهش

در سال ۱۹۹۰ در (میلر و همکاران، ۱۹۹۰) یک پایگاه‌داده واژگانی برای زبان انگلیسی به نام وردنت (میلر، ۱۹۹۵) (فلباوم،

---

[73] WordNet
[74] Synset

۱۹۹۸) ارائه کردند که کلمات انگلیسی را به مجموعه‌های هم‌معنی، به نام هم‌نشیم گروه‌بندی می‌نماید و از هر کلمه یک تعریف مختصر و چند مثال از کاربرد آن نیز وجود دارد و ضمناً بعضی از روابط بین این هم‌نشیم‌ها و اعضای آن‌ها را نیز در خود دارد. پس از ارائه وردنت، با الگو قراردادن ساختار وردنت، پایگاه‌داده‌های واژگانی برای زبان‌های دیگر نیز ارائه شد (بوند و پایک، ۲۰۱۲) (ووسن، ۱۹۹۸) (ووسن، ۲۰۰۴)، که همانند وردنت هم‌نشیم‌های زبان خود را به طور نظام‌مند گروه‌بندی می‌کردند. ما به مجموعه این پایگاه‌داده‌های واژگانی «پایگاه‌داده واژگانی شبه‌وردنت[۷۵]» (WLD) می‌گوییم. WLDها کاربردهای گسترده‌ای در پردازش زبان‌های طبیعی (وی و ژاوو، ۲۰۱۵) (رفورجاتو و همکاران، ۲۰۱۵) مهندسی دانش (یان و ونگ، ۲۰۱۸) (لواسیک و پوهار، ۲۰۱۶) و مهندسی هستان‌شناسی (مدالی و سولوچانا، ۲۰۱۶) (بیمسان و هول، ۲۰۱۶) و تعبیه هستان‌شناسی (سعیدی و همکاران ۲۰۱۹)(بن‌عرب و همکاران ۲۰۲۳) دارند.

در WLD ها، همه اعضای یک هم‌نشیم به یک میزان عضو یک هم‌نشیم در نظر گرفته می‌شوند، یعنی همه اعضا در یک سطح معنای آن هم‌نشیم را نشان می‌دهند؛ به‌عبارت‌دیگر، WLDها فرض می‌کنند که هم‌نشیم مجموعه‌های قطعی (غیرفازی) هستند؛ اما این فرض ساده همیشه نمی‌تواند پیچیدگی ذاتی «معنا» در متن کاوی را مدل کند. به‌عنوان‌مثال، هم‌نشیم زیر را از وردنت در نظر بگیرید:

(‘flower.n.02’): {flower, bloom, blossom};

این مجموعه شامل کلماتی است که به‌صورت بالقوه (یکی از واژه‌بارها) به‌عنوان یکی از معانی کلمه به این مفهوم به کار می‌رود: «عضو تولیدمثل گیاهان گل‌دار که به‌صورت خاص نمایان و رنگی است» (در وردنت تعریف هر هم‌نشیم به‌صورت توصیفی ارائه‌شده است.)

پیش از ادامه این بحث، لازم است که مفهوم «بن‌واژه[۷۶]» و مفهوم «واژه‌بار» را در WLDها معرفی کنیم: هر کلمه بدون توجه به معانی متفاوتی که می‌تواند داشته باشد «بن‌واژه» نامیده می‌شود. برای مثال "bloom" مستقل از معنایی که می‌تواند داشته باشد، یک بن‌واژه در نظر گرفته می‌شود. این برای همه کلمات لغت‌نامه برقرار است. علاوه بر این، یک معنای خاص یک بن‌واژه که قاعدتاً عضوی از یک هم‌نشیم خاص است، «واژه‌بار» گفته می‌شود. برای مثال هم‌نشیم ذکرشده در بالا برای کلمه "bloom" (شکوفه) یک واژه‌بار گفته می‌شود. هر بن‌واژه می‌تواند چند واژه‌بار داشته باشد؛ به‌عبارت‌دیگر هر بن‌واژه می‌تواند عضو بیش از یک هم‌نشیم باشد.

همان‌طور که در بالا گفته شد، معمولاً واژه‌بارهای مربوط به یک هم‌نشیم به یک میزان معنای آن هم‌نشیم را نشان نمی‌دهند. به‌عنوان‌مثال هم‌نشیم (‘flower.n.02’) که هم‌نشیم مربوط به دومین معنای اسمی بن‌واژه flower است (flower و bloom و ...) به طور ۱۰۰٪ بر تعریف «عضو تولیدمثل گیاهان گل‌دار که به‌صورت خاص نمایان و رنگی است» منطبق نیستند. برای حل این مشکل مفهوم هم‌نشیم فازی پیشنهاد شد. از سال ۲۰۰۵ مطالعاتی در این زمینه انجام‌شده است که در آن‌ها هم‌نشیم، یک مجموعه فازی در نظر گرفته می‌شود.

در (ولدال، ۲۰۰۵) بدون استفاده از اصطلاح هم‌نشیم فازی (حتی بدون استفاده از اصطلاح «هم‌نشیم»)، الگوریتمی

[75] WordNet-like Lexical Databases
[76] Lemma

برای ایجاد خوشه‌های معنایی فازی پیشنهاد کرد که این خوشه‌های معنایی همان هم‌نشیم‌ها بودند. او الگوریتم خود را روی زبان نروژی پیاده نمود.

در (بورین و فورسبرگ، ۲۰۱۰) برای اولین‌بار اصطلاح «هم‌نشیم فازی[۷۷]» را تعریف کردند. آن‌ها این مفهوم را از دیدگاه زبان‌شناسی بررسی کردند و آن را بر اساس مفهوم «اجتناب از هم‌معنایی[۷۸]» (هارفورد، ۲۰۰۳) ایجاد کردند. اصل «اجتناب از هم‌معنایی» به این نکته اشاره دارد که در زبان طبیعی، به‌ندرت دو کلمه کاملاً هم‌معنی هستند. یعنی بین معنای دو کلمه در یک زبان، حتماً تفاوت وجود دارد.

در این مطالعات (بورین و فورسبرگ، ۲۰۱۰) (بورین و فورسبرگ، ۲۰۱۰) از Synlex (که یک پیکره لغات مترادف ساخته‌شده توسط مردم است (کان و روسل، ۲۰۰۵) که شامل درجه ترادف زوج لغات است و به‌وسیله جمع‌سپاری ایجاد شده است) و SALDO (بورین، ۲۰۰۵) (بورین و فورسبرگ، ۲۰۰۹) (یک پیکره جامع سوئدی به‌صورت نحوی معنایی است که روابط انجمنی بین احساس کلمات و عبارات را بیان می‌کند) استفاده کردند؛ و با استفاده از آن برای ایجاد هم‌نشیم‌های فازی به زبان سوئدی الگوریتمی ارائه نمودند. در سال ۲۰۱۱ (الیورا و گومز، ۲۰۱۱) دومین کسانی بودند که از منظر زبان‌شناسی در مورد هم‌نشیم‌های فازی بررسی انجام دادند. در این مطالعه، آن‌ها الگوریتم خود را در زبان پرتغالی پیاده کردند و یک هم‌نشیم فازی به زبان پرتغالی ارائه کردند.

اما هیچ‌کدام از هم‌نشیم‌های فازی ارائه‌شده فوق، برای کاربردهای واقعی که از وردنت‌های موجود استفاده می‌کنند، مناسب نیستند؛ چرا که در مطالعات ذکرشده، یا هم‌نشیم‌ها از پیش تعریف شده نیستند و بعد از اجرای الگوریتم تعیین می‌شوند (هم‌نشیم‌های فازی خروجی خوشه‌بندی هستند (ولدا، ۲۰۰۵) (الیورا و گومز، ۲۰۱۱)) یا اینکه از یک پایگاه‌داده لغوی موجود استخراج شده‌اند (مانند SALDO در (بورین و فورسبرگ، ۲۰۱۰) که یک شبه‌وردنت نیست) اما خروجی الگوریتم یک پایگاه‌داده لغوی اصلاح شده است و بنابراین هم‌نشیم‌های آن نسخه فازیِ هم‌نشیم‌های قبلی نیست.

این در حالی است که (همان‌طور که ذکر شد) WLDها توجه و کاربرد گسترده‌ای را به خود اختصاص داده‌اند و کاربردهای آن‌ها بر اساس WLDهای موجود است؛ بنابراین ایجاد هم‌نشیم‌های متفاوت با اعضای WLDهای موجود، به معنی نادیده‌گرفتن طیف گسترده‌ای از کاربردهای WLD است.

به همین منظور، در (حسینی و همکاران، ۲۰۱۶) ایدهٔ ایجاد یک الگوریتم برای فازی کردن هم‌نشیم‌های قطعی در WLDهای موجود ارائه کردند، ولی به ایدهٔ کلی ارائه شده بسنده کردند.

در بخش آتی، به‌عنوان مقدمه، با الهام‌گرفتن از ایدهٔ کلی مطرح شده در (حسینی و همکاران، ۲۰۱۶) یک الگوریتم مستقل از زبان ارائه می‌دهیم که قادر است WLDهای موجود را فازی نماید. این الگوریتم می‌تواند یک تابع عضویت برای هم‌نشیم‌های از پیش تعیین شده در هر زبانی تخصیص دهد. برای اجرای این الگوریتم نیاز به یک پیکره بزرگ از متون آن زبان و یک سامانه رفع ابهام واژه‌بار[۷۹] (WSD) به‌عنوان ورودی الگوریتم داریم. به این منظور الگوریتم خود را روی زبان

[77] Fuzzy synset
[78] Synonymy avoidance
[79] Word Sense Disambiguation

انگلیسی پیاده کرده و با استفاده از پیکره باز ملی آمریکا[80] (OANC) و سامانه رفع ابهام گراف‌محور و شناخته‌شده UKB، نسخه فازی وردنت انگلیسی ایجاد می‌نماییم. جزئیات این الگوریتم در بخش بعد ذکر خواهد شد.

شایان‌ذکر است در فصل‌های بعد، برای اعتبارسنجی[81] هم‌نشیم‌های فازی به‌دست‌آمده، ابتدا سبد هم‌نشیم‌های فازی (BoFS) و سبد واژه‌بارهای فازی (BoFWS) را به‌عنوان دو مدل کمکی متن‌کاوی معرفی کرده، سپس با بهره‌گیری از این مدل‌ها در فرایند تحلیل احساسات بررسی می‌کنیم که تا چه حد نتایج بهبود می‌یابد؛ تا با این کار با اثبات کیفیت بالاتر نتایج حاصل از BoFS و BoFWS نسبت به نسخهٔ غیرفازی آنها، کیفیت هم‌نشیم‌های فازی ساخته‌شده برای زبان انگلیسی را تأیید نماییم. مدل‌های مذکور در بخش سوم و آزمایش‌های تحلیل احساسات یادشده در بخش آخر توضیح داده خواهند شد.

۳) تولید هم‌نشیم‌های فازی برای هم‌نشیم‌های از پیش تعیین شده

در این بخش، الگوریتم ایجاد وردنت فازی (که در بالا به آن اشاره شد) را شرح خواهیم داد.

این الگوریتم به ورودی‌های زیر نیاز دارد:

یک پیکره بزرگ از اسناد در زبان موردنظر

یک سامانه رفع ابهام واژه‌بار (WSD)

الگوریتم از ۳ مرحله زیر تشکیل شده است:

ابتدا تمام پیکره متنی را به‌وسیله سامانه WSD رفع ابهام می‌کنیم تا تعداد تکرار واژه‌بارها از هر هم‌نشیم را بشمارد. در این مرحله برای هر هم‌نشیم یک نمودار بسامد از تعداد تکرارهای اعضای آن (واژه‌بارها) رسم می‌کنیم.

۱. نمودارهای بسامد را با نرمال کردن آن‌ها به نمودار احتمال تبدیل می‌کنیم. فرض کنید X یک متغیر تصادفی از فضای حالت واژه‌بارهای U باشد. مقدار زیر برابر است با «احتمال تساوی X با واژه‌بار ω، با این فرض که هم‌نشیم x برابر با S باشد» (البته اگر اندازه پیکره به‌اندازه کافی بزرگ باشد)

$$p(\omega) = \frac{\text{frequency of } \omega}{\text{frequency of all of the word-senses of S}}$$ رابطه ۱)

۲. تابع توزیع احتمال واژه‌بارها را به‌وسیله تابع تبدیل احتمال به امکان به تابع توزیع امکان تبدیل می‌کنیم.

$$1983: \pi(c_i) = \sum_{c_j \in C} \min\left(p(c_j), p(c_i)\right)$$ رابطه ۲)

$$1993: \pi(c_i) = \sum_{c_j | p(c_j) \le p(c_i)} p(c_j)$$ رابطه ۳)

برای اثبات معنادار بودن خروجی الگوریتم، بر اساس تعریف ریاضی «امکان» که در فصل گذشته ذکر شد، U را مجموعه همه واژه‌بارها قرار می‌دهیم و F را نشان‌دهنده یک هم‌نشیم فازی در نظر می‌گیریم که زیرمجموعه‌ای از جهان-

[80] Open American National Corpus
[81] validation

گفتمان U است و توسط تابع عضویت آن $\mu_F(u)$، تعیین می‌شود که از آن به‌عنوان میزان انطباق واژه‌بار u با هم‌نشیم F تعبیر می‌شود.

سپس، X را یک مقدار متغیر در فضایی از واژه‌بارهای U قرار می‌دهیم و فرض می‌کنیم رفتار هم‌نشیم F در محدودیت فازی R(X) با X انجمن شده[82] باشد.

اینک، گزاره «مقدار X برابر F اسـت»، به R (X) = F ترجمه می‌شـود وقتی که R(X) به‌عنوان «هم‌نشـیم متعلق به X» تعبیر می‌شود.

گزاره «مقدار X برابر F است» توزیع امکان Π_X را به X منطبق می‌کند که مساوی R(X) فرض می‌شود (یعنی $\Pi_X = R(X)$). به این ترتیب، تابع توزیع امکان Π_X (توسط π_X مشخص‌شده) به‌صورت عددی برابر با تابع عضویت F تعریف می‌شود (یعنی $\pi_X \triangleq \mu_F$).

بنابراین $\pi_X(u)$ برابر است با امکان اینکه X برابر با واژه‌بار u باشد (یعنی امکان X = u)، با توجه به اینکه «مقدار X برابر F است» (یا به عبارتی R (X) = F یا «هم‌نشیم مطابق با X برابر F است»)، $\pi_X(u)$ برابر با درجه عضویت واژه‌بار u در هم‌نشیم F تعریف می‌شود.

پس نتیجه می‌گیریم که خروجی مرحله ۳ الگوریتم، بر اساس تعریف امکان (که در فصل ۱ ذکر شد)، برابر با درجه عضویت هم‌نشیم مربوطه (S) می‌باشد.

در زیر، شبه کد الگوریتم ذکر شده را می‌بینیم.

```
WSF[][] = WordSenseDisambiguation(corpus);
//WSF is Words-Sense Frequency matrix. The 1st dimension is for synsets and the 2nd dimension is
    for its word-senses. Each cell represents the frequency of a synset's member (word-sense)
    in the whole corpus.
For i = 1 to total number of synsets
        synSize = numberOfWordSenses(synset[i]);
        totalFrequencyOfSynset = 0;
        For j = 1 to synSize
                totalFrequencyOfSynset += WSF[i][j];
        For j = 1 to synSize
                PMV[i][j] = WSF[i][j] / totalFrequencyOfSynset;
                //PMV is Probability Mass Value matrix; the dimensions are the same as what in
                    WSF.
        For j = 1 to synSize
                possibility1983OfJ = 0;
                possibility1993OfJ = 0;
                pIJ = PMV[i][j];
                For m = 1 to synSize
                        pIM = PMV[i][m];
                        possibility1983OfJ += min(pIJ,pIM);
                        possibility1993OfJ += piecewise(pIM <= pIJ , pIM , 0);
                Possibility1983[i][j] = possibility1983OfJ;
```

[82] Associated

```
Possibility1993[i][j] = possibility1993OfJ;
```

اگرچه کارایی این الگوریتم به‌صورت نظری نشان‌داده‌شده است، در ادامه ابزار متن‌کاوی لازم برای ارزیابی تجربی و آزمایش‌های تجربی را برای نشان‌دادن کارایی عملی وردنت‌های فازی شده ارائه می‌نماییم.

ضمناً، الگوریتم ارائه شده را بر روی وردنت استاندارد (همان‌گونه که پیش‌تر اشاره شد) با استفاده از پیکره باز ملی آمریکا و سامانه ابهام‌زدای UKB اعمال شده است و نتایج از طریق پیوند http://dmls.iust.ac.ir/CogLing/FWN.zip قابل‌دسترسی می‌باشد.

۴) برخی از مدل‌های متن‌کاوی موجود

در این قسمت برخی از مدل‌های متن‌کاوی موجود که بر اساس آنها کارهای پردازش متن بسیاری انجام شده است را ذکر می‌کنم. سپس در قسمت بعد مدل‌های متن‌کاوی خود را شرح می‌دهیم.

۴–۱) سبد کلمات[۸۳] (BoW)

ساده‌ترین و شناخته‌شده‌ترین مدل برای متن‌کاوی، مدل unigram(مدل n-gram وقتی n=1) است؛ که این مدل به‌عنوان سبد کلمات شناخته می‌شود (ژانگ و ژائو، ۲۰۱۰) . در این مدل پس از گرفتن یک سند متنی؛ مانند یک توییت یا... در مدل سبد کلمات ابتدا ایست‌واژه حذف می‌شوند. سپس هر کلمه به‌عنوان یک ویژگی در نظر گرفته می‌شود. برای مثال در توییت زیر:

This happy boy is happy because he is not a sad boy! Isn’t the boy? The boy is!

مدل BoW ابتدا علامت‌ها و ایست‌واژه‌ها را حذف می‌کند. (به‌عنوان‌مثال:“because” “is” “This” و ...) سپس کلمات باقی‌مانده را به‌عنوان ویژگی‌های سند در نظر می‌گیرد؛ و درنهایت خروجی زیر را تولید می‌کند:

{(Happy, 2), (boy, 4), (sad, 1)}.

۴–۲) سبد هم‌نشیم‌ها[۸۴] (BoS)

سبد هم‌نشیم‌ها یک توسعه از سبد کلمات (BoW) است که بجای درنظرگرفتن یک کلمه، هم‌نشیم مرتبط با آن کلمه را به‌عنوان ویژگی در نظر می‌گیرد.

در (والی و اسلم، ۱۹۹۹) برای اولین‌بار این مدل را پیشنهاد دادند. هرچند آن را «سبد مفاهیم» نامیدند و هر هم‌نشیم را یک مفهوم در نظر گرفتند.

بااین‌وجود در (مانجولا و اقیلا، ۲۰۰۳) برای اولین‌بار اصطلاح مشابه سبد هم‌نشیم‌ها را پیشنهاد کردند یعنی «سبد هم‌نشیم‌های کلمات». این ایده در سال‌های بعد در تحقیقات شناخته‌شده‌تری در این شاخه ادامه پیدا کرد؛ مانند مطالعات (سمرارو و لوپز، ۲۰۰۵)، (یه و بلدوین، ۲۰۰۶)، (باسیل و همکاران، ۲۰۰۷)، (لوپز و همکاران، ۲۰۰۷)و (سمرارو و همکاران،

[83] Bag of Words
[84] Bag of synsets

2007)، (گمیس و همکاران، ۲۰۰۸)، و یکی از کاربردی‌ترین مطالعات در این زمینه SentiWordNet (SWN) 3.0 در (باسیانلا و همکاران، ۲۰۱۰)بود که یک ایده مشابه مرحله ۲ الگوریتم ساخت SWN استفاده کرده است. این ایده در تحقیقات مربوط به متن‌کاوی در زبان‌های دیگر نیز بکار گرفته‌شده (لوپز و همکاران، ۲۰۱۰) (پورامینی و مینایی، ۲۰۱۶). بااین‌وجود در سال‌های اخیر، توجه به این مدل کم شده است.

برای فهم بهتر مدل BoS به ذکر یک مثال اکتفا می‌کنیم:

هم‌نشیم‌های جدول ۱ را در نظر بگیرید. مدل BoS مربوط به توییت زیر:

I think that we are supposing this plant incorrectly as a flower; only because this part is like bloom/blossom.

مجموعه زیر خواهد بود:

{(S1, 2), (S2, 1), (S3, 3), (S4, 1), (S5, 1), (S6, 1)}.

جدول ۱: فهرست چند هم‌نشیم با Synset ID دلخواه

Synset ID	Word senses
S1	think, opine, suppose, imagine, reckon, guess
S2	plant, flora, plant life
S3	flower, bloom, blossom
S4	falsely, incorrectly
S5	merely, simply, just, only, but
S6	part, portion

۵) مدل‌های متن‌کاوی ایجادشده

در این بخش، چند مدل جدید فازی برای متن‌کاوی، بر اساس هم‌نشیم‌های پیشنهادی در فصل قبلی پیشنهاد می‌کنیم.

۵–۱) سبد هم‌نشیم‌های فازی[۸۵] (BOFS)

در بخشی از این پژوهش، ما مدل متنی‌ای به نام سبد هم‌نشیم‌های فازی یا BOFS ارائه دادیم که به‌عنوان مدل فازی شده‌ای از مدل BOS محسوب می‌شود. اگرچه با توجه به اینکه در فصل‌های گذشته هم‌نشیم‌های فازی به طور کامل معرفی شده است، مدل BOFS به‌سادگی قابل‌درک است. بااین‌حال برای روشن‌شدن بیشتر، مثال ارائه‌شده در بخش BOS را در این مدل بررسی می‌کنیم. جدول ۳–۲ را به‌عنوان نسخه اصلاح‌شده جدول ۲ در نظر بگیرید که در آن درجه عضویت هر واژه مشخص شده است.

جدول ۲: فهرست چند هم‌نشیم با Synset ID دلخواه.

Synset ID	Word senses
S1	think (1), opine (0.7), suppose (0.8), imagine (0.4), reckon (0.2), guess (0.5)
S2	plant (1), flora (0.8), plant life (0.7)
S3	flower (1), bloom (0.7), blossom (0.6)
S4	falsely (1), incorrectly (0.9)
S5	merely (1), simply (0.5), just (1), only (1), but (0.8)
S6	part (1), portion (0.7)

[85] Bag of Fuzzy Synsets

اکنون برای مدل‌سازی مثال ارائه‌شده در بخش BoS بجای اضافه کردن تعداد حضور در هم‌نشیم، میزان عضویت هر واژه‌بار را اضافه می‌کنیم؛ بنابراین مثال مذکور به این صورت مدل می‌شود:

{(S1, 1.8), (S2, 1), (S3, 2.3), (S4, 0.9), (S5, 1), (S6, 1)}.

۵-۲) سبد واژه‌بارهای فازی[86] (BoFWS)

با استفاده از یک ایده مشابه به آنچه که در BoFS پیشنهاد شد، یک مدل متن‌کاوی جدید با استفاده از واژه‌بارها، به‌جای همنشیم‌ها، به نام سبد واژه‌بارهای فازی(BoFWS) تعریف می‌نماییم. این مدل را می‌توان به‌عنوان نسخه فازی شده از سبد واژه‌بارها (BoWS) در نظر گرفت که در ابتدا در سال ۱۹۹۵ ارائه شد (اسمیتون، ۱۹۹۵). اگرچه BoWS یک مدل متن‌کاوی پرطرف‌دار نیست، ما معتقدیم که این نقص این مدل، به این دلیل است که به همه واژه‌بارها یک همنشیم وزن‌های مشابه اختصاص می‌دهد؛ و همچنین ادعا می‌کنیم که اگر این مدل فازی شود، این مشکل را برطرف می‌کند و باعث بالارفتن دقت BoWS می‌شود.

برای ذکر مثال از همان مثال قبلی که در بالا توضیح داده شده استفاده می‌کنیم و BoFWS را برای آن نشان می‌دهیم.

{(WS(think.v.02.think), 1), (WS(think.v.02.suppose), 1.6), (WS(plant.n.02.plant), 2), (WS(flower.n.02.flower), 1), (WS(flower.n.02.bloom), 0.7), (WS(flower.n.02.blossom), 0.6), (WS(falsely.r.02.incorrectly), 1.8) , (WS(merely.r.01.only), 1) , (WS(part.n.02.part), 1)}.

همان‌طور که می‌بینید، واژه‌بارها به‌عنوان ویژگی‌های متن در نظر گرفته می‌شوند (مانند BoWS) ، اما بسامد هر واژه‌بار در درجه عضویت آن ضرب می‌شود. به‌عنوان‌مثال، بسامد واژه‌بار *WS(falsely.r.02.incorrectly)* برابر ۲ است؛ اما آن را در ۰.۹ ضرب می‌کنیم و ۱.۸ را به‌عنوان مقدار این ویژگی در نظر می‌گیریم.

۶) آزمایش‌ها

در این بخش به بیان برخی از آزمایش‌ها صورت‌گرفته جهت ارزیابی مدل پیشنهادی می‌پردازیم.

جهت ارزیابی وردنت فازی ایجاد شده، به حل یک مسئله تحلیل احساسات می‌پردازیم. ازآنجایی‌که مسئله تحلیل احساسات در سطح سند را می‌توان نوعی مسئله رده‌بندی اسناد متنی محسوب کرد، برای ارزیابی مدل خود با استفاده از مدل زبانی معرفی شده BOFS یک رده‌بند ایجاد می‌کنیم و به‌وسیله آن مسئله تحلیل احساسات را حل می‌کنیم.

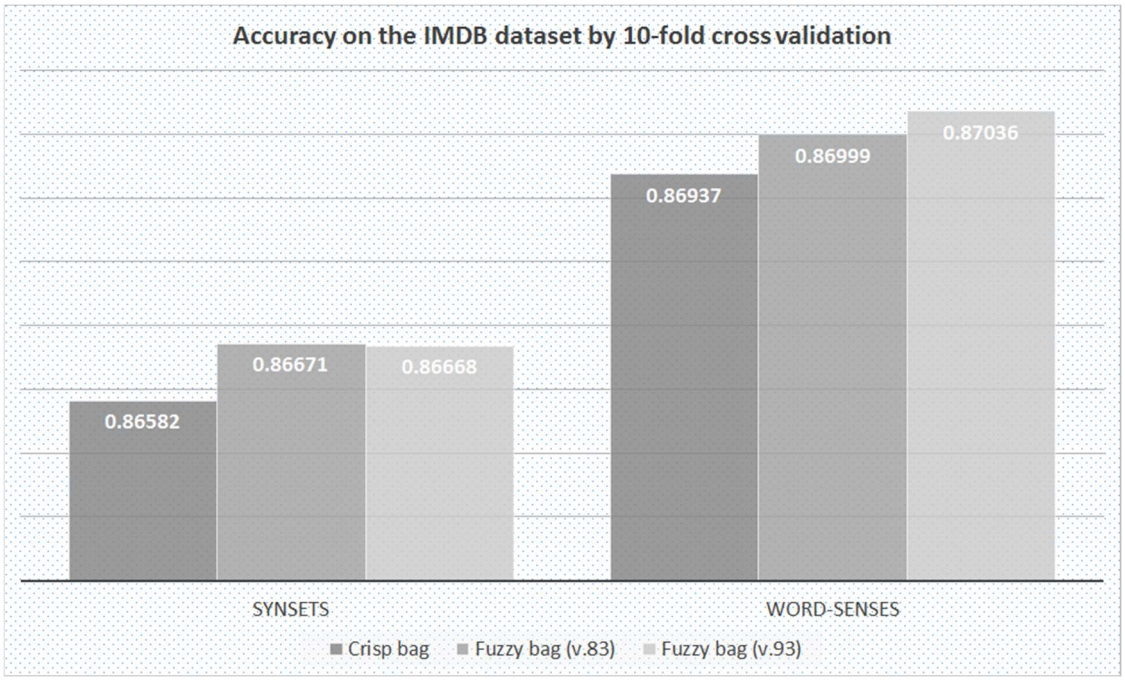
این آزمایشات بر روی یک پیکره بزرگ زبانی شامل ۵۰ هزار سند تحلیلی است که در دو رده مثبت و منفی در سطح سند برچسب گذاری شده اند (ماس و همکاران، ۲۰۱۱). جهت رده بندی اسناد از رده‌بند NB-SVM[87] استفاده شد که در سال ۲۰۱۲ برای رده بندی متون ارائه شد. پژوهشگران مطرحی از جمله منینگ (ونگ و منینگ، ۲۰۱۱) به عنوان یک رویکرد سریع و در عین حال با کیفیت برای حل مسئله فوق از الگوریتم NBSVM (که یک ماشین بردار پشتیبان بیزی ساده است) استفاده کرده‌اند. ما نیز در در این پژوهش از الگوریتم NBSVM استفاده می‌کنیم.

[86] Bag of Fuzzy Word-Senses
[87] Naïve Bayes Support Vector Machine

۶-۱) نتیجه آزمایش

جهت مقایسه خروجی روش پیشنهادی با روش‌های قبلی، ابتدا با استفاده از مدل زبانی BOS و BOWS یک رده بنده بوسیله NBSVM ساختیم و این دو را ارزیابی کردیم. داده‌ها به دو دسته ارزیابی و آموزش تقسیم شدند که ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد دیگر برای ارزیابی خروجی استفاده شد. همچنین از اعتبارسنجی متقابل به صورت ۱۰-fold استفاده شد.

سپس با استفاده از هر یک از مدل‌های BOFS و BOFWS دو رده‌بند به‌وسیله تابع‌های تبدیل احتمال به امکان ۸۳ (رابطه ۲) و ۹۳ (رابطه ۳) ایجاد کردیم و خروجی‌های این دو رده‌بند را با خروجی‌های رده‌بندهای قطعی خود مقایسه کردیم. در (شکل ۱) خروجی‌های رده‌بند نمایش‌داده‌شده است.



شکل ۱: مقایسه BoS، BOFS، BoWS و BoFWS به‌وسیله نسخه‌های مختلف فازی

جهت محاسبه درصد اطمینان مقایسه انجام شده، از روش آزمون آماری فرض صفر استفاده کردیم. جدول (۳) مقادیر p-values مربوط به فرض صفر است که نتایج مربوط به مقایسه نسخه‌های فازی با نتایج نسخه‌های قطعی است.

جدول ۳: ارزیابی مقدار p در آزمون دانشجو برای مقایسهٔ معناداری تفاوت نتایج BoFS و BoFWS، نسخه‌های فازی ۱۹۸۳ و ۱۹۹۳

p-value Table	Equality of fuzzy v.93 results and their crisp version results	Equality of fuzzy v.83 results and their crisp version results
Synset	0.015	0.004
Word-Sense	0.003	0.019

همان‌طور که در جدول ۱ مشاهده می‌کنید، تمام نتایج حاصل از p-values کاملاً کوچک هستند و ۹۸٪ اطمینان به برتر بودن نتایج حاصل از تحلیل احساسات به‌وسیله وردنت فازی نسبت به نسخه قطعی آن، می‌دهد. به طور خاص، بهترین نتیجه متعلق به BoFWS v.93 است که اطمینان ۹۹.۷٪ به برتری آن می‌دهد.

۷) منابع و مراجع

Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, No. 2010, pp. 2200-2204). DOI:10.1016/S0165-0114(98)00137-7

Basile, P., Degemmis, M., Gentile, A. L., Lops, P., & Semeraro, G. (2007). The jigsaw algorithm for word sense disambiguation and semantic indexing of documents. In *AI* IA 2007: Artificial Intelligence and Human-Oriented Computing: 10th Congress of the Italian Association for Artificial Intelligence, Rome, Italy, September 10-13, 2007. Proceedings 10* (pp. 314-325). Springer Berlin Heidelberg. DOI:10.1016/S0165-0114(98)00137-7

Benarab, A., Sun, J., Rafique, F., & Refoufi, A. (2023). Global Ontology Entities Embeddings. *IEEE Transactions on Knowledge and Data Engineering*. DOI:10.1016/S0165-0114(98)00137-7

Bender, E. M., & Koller, A. (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. *In Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5185-5198). DOI:10.1016/S0165-0114(98)00137-7

Bimson, K. D., Hull, R. D., & Nieten, D. (2016). The lexical bridge: A methodology for bridging the semantic gaps between a natural language and an ontology. *Semantic Web: Implications for Technologies and Business Practices*, 137-151. DOI:10.1016/S0165-0114(98)00137-7

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. DOI:10.1016/S0165-0114(98)00137-7

Bond, F., & Paik, K. (2012, January). A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)* (pp. 64-71). DOI:10.1016/S0165-0114(98)00137-7

Borin, L. (2005). Mannen är faderns mormor: Svenskt associationslexikon reinkarnerat. *LexicoNordica*, (12). DOI:10.1016/S0165-0114(98)00137-7

Borin, L., & Forsberg, M. (2009, May). All in the family: A comparison of SALDO and WordNet. In Proceedings of the Nodalida 2009 Workshop on WordNets and other Lexical Semantic Resources–between Lexical Semantics, Lexicography, Terminology and Formal Ontologies (pp. 7-12). DOI:10.1016/S0165-0114(98)00137-7

Borin, L., & Forsberg, M. (2010). Beyond the synset: Swesaurus–a fuzzy Swedish wordnet. In *Workshop on Re-thinking synonymy: Semantic sameness and similarity in languages and their description. Helsinki*. DOI:10.1016/S0165-0114(98)00137-7

Borin, L., & Forsberg, M. (2010, May). From the people's synonym dictionary to fuzzy synsets-first steps. In *Proceedings of the LREC 2010 workshop Semantic relations. Theory and Applications* (pp. 18-25). DOI:10.1016/S0165-0114(98)00137-7

Chu, J. S., & Evans, J. A. (2021). Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences*, *118*(41), e2021636118. DOI:10.1016/S0165-0114(98)00137-7

De Gemmis, M., Lops, P., Semeraro, G., & Basile, P. (2008, October). Integrating tags in a semantic content-based recommender. In *Proceedings of the 2008 ACM conference on Recommender systems* (pp. 163-170). DOI:10.1016/S0165-0114(98)00137-7

Hajiali, M. (2020). Big data and sentiment analysis: A comprehensive and systematic literature review. *Concurrency and Computation: Practice and Experience*, 32(14), e5671. DOI:10.1016/S0165-0114(98)00137-7

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1), 1. DOI:10.1016/S0165-0114(98)00137-7

Hershcovich, D., & Donatelli, L. (2021). It's the meaning that counts: the state of the art in NLP and semantics. *KI-Künstliche Intelligenz*, 35(3-4), 255-270. DOI:10.1016/S0165-0114(98)00137-7

Hossayni, S. A., Akbarzadeh-T, M. R., Recupero, D. R., Gangemi, A., & Josep Lluís de la Rosa i Esteva. (2016). Fuzzy Synsets, and Lexicon-Based Sentiment Analysis. In *EMSA-RMed@ ESWC*. DOI:10.1016/S0165-0114(98)00137-7

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In *Advances in Artificial Life: 7th European Conference, ECAL 2003, Dortmund, Germany, September 14-17, 2003. Proceedings 7* (pp. 442-451). Springer Berlin Heidelberg. DOI:10.1016/S0165-0114(98)00137-7

Ivasic-Kos, M., Pobar, M., & Ribaric, S. (2016). Two-tier image annotation model based on a multi-label classifier and fuzzy-knowledge representation scheme. *Pattern recognition*, *52*, 287-305. DOI:10.1016/S0165-0114(98)00137-7

Kann, V., & Rosell, M. (2006, May). Free construction of a free Swedish dictionary of synonyms. In *Proceedings of the 15th Nordic Conference of Computational Linguistics (NODALIDA 2005)* (pp. 105-110). DOI:10.1016/S0165-114(98)00137-7

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: Sta te of the art, current trends and challenges. *Multimedia tools and applications*, 82(3), 3713-3744. DOI:10.1016/S0165-0114(98)00137-7

Lops, P., Degemmis, M., & Semeraro, G. (2007). Improving social filtering techniques through wordnet-based user profiles. In *User Modeling 2007: 11th International Conference, UM 2007, Corfu, Greece, July 25-29, 2007. Proceedings 11* (pp. 268-277). Springer Berlin Heidelberg. DOI:10.1016/S0165-0114(98)00137-7

Lops, P., Musto, C., Narducci, F., De Gemmis, M., Basile, P., & Semeraro, G. (2010, September). Mars: a multilanguage recommender system. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems* (pp. 24-31). DOI:10.1016/S0165-0114(98)00137-7

Madalli, D., Sulochana, A., & Singh, A. K. (2016). COMAT: core ontology of matter. *Program*, *50*(1), 103-117. DOI:10.1016/S0165-0114(98)00137-7

Manjula, D., Aghila, G., & Geetha, T. V. (2003, April). Document knowledge representation using description logics for information extraction and querying. In *Proceedings ITCC 2003. International conference on information technology: Coding and computing* (pp. 189-193). IEEE. DOI:10.1016/S0165-0114(98)00137-7

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39-41. DOI:10.1016/S0165-0114(98)00137-7

Miller, G. A. (1998). *WordNet: An electronic lexical database*. MIT press. DOI:10.1016/S0165-0114(98)00137-7

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, *3*(4), 235-244. DOI:10.1088/S0165-0114(98)32014-7

Oliveira, H. G., & Gomes, P. (2011, June). Automatic discovery of fuzzy synsets from dictionary definitions. In *Twenty-Second International Joint Conference on Artificial Intelligence*. DOI:10.2015/S0165-0114(98)00362-7

Pouramini, J., & Minaei-Bidgoli, B. (2016). A New Synthetic Oversampling Method Using Ontology and Feature Selection in Order to Improve Imbalanced Textual Data Classification in Persian Texts. *Bulletin de la Société Royale des Sciences de Liège*, *85*, 358-375. DOI:10.2014/S0165-0114(98)326001-7

Reforgiato Recupero, D., Presutti, V., Consoli, S., Gangemi, A., & Nuzzolese, A. G. (2015). Sentilo: frame-based sentiment analysis. *Cognitive Computation*, *7*, 211-225. DOI:10.1016/S0165-0114(98)00137-7

Saedi, C., Branco, A., Rodrigues, J., & Silva, J. (2018, July). Wordnet embeddings. *In Proceedings of the third workshop on representation learning for NLP* (pp. 122-131). DOI:10.2102/S0165-0114(98)00136-7

Semeraro, G., Degemmis, M., Lops, P., & Basile, P. (2007, January). Combining Learning and Word Sense Disambiguation for Intelligent User Profiling. In *IJCAI* (Vol. 7, pp. 2856-2861). DOI:10.3021/S0165-0114(98)21500-7

Semeraro, G., Lops, P., & Degemmis, M. (2005, November). WordNet-based user profiles for neighborhood formation in hybrid recommender systems. In *Fifth International Conference on Hybrid Intelligent Systems (HIS'05)* (pp. 6-pp). IEEE. DOI:10.1016/S0165-0114(98)00137-7

Sharma, S., & Jain, A. (2020). Role of sentiment analysis in social media security and analytics. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery*, 10(5), e1366. DOI:10.3265/S0165-0114(98)032154-7

Smeaton, A. F., Kelledy, F., O'Donnell, R., Quigley, I., Richardson, R., & Townsend, E. (1995, October). Low Level Language Processing for Large Scale Information Retrieval: What Techniques Actually Work. In *Proceedings of a Workshop: Terminology, Information Retrieval, and Linguistics*. DOI:10.2158/S0165-0114(98)036214-7

Velldal, E. (2005). A fuzzy clustering approach to word sense discrimination. In *Proceedings of the 7th International conference on Terminology and Knowledge Engineering* (pp. 279-292). DOI:10.1025/S0165-0114(98)00362-7

Vossen, P. (1998). Introduction to eurowordnet. *EuroWordNet: A multilingual database with lexical semantic networks*, 1-17. DOI:10.0125/S0165-0114(98)326517-7

Vossen, P. (2004). Eurowordnet: a multilingual database of autonomous and language-specific wordnets connected via an inter-lingualindex. *international journal of Lexicography*, *17*(2), 161-173. DOI:10.0110/S0165-0114(98)625145-7

Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A semantic approach for text clustering using WordNet and lexical chains. *Expert Systems with Applications*, *42*(4), 2264-2275. DOI:10.2158/S0165-0114(98)326914-7

Whaley, J. M. (1999). An application of word sense disambiguation to information retrieval. *Dartmouth College, Department of Computer Science*. DOI:10.1016/S0165-0114(98)021574-7

Yan, J., Wang, C., Cheng, W., Gao, M., & Zhou, A. (2018). A retrospective of knowledge graphs. *Frontiers of Computer Science*, *12*, 55-74. DOI:10.6574/S0165-0114(98)00654-7

Ye, P., & Baldwin, T. (2006, November). Verb sense disambiguation using selectional preferences extracted with a state-of-the-art semantic role labeler. In *Proceedings of the Australasian Language Technology Workshop 2006* (pp. 139-148). DOI:10.1857/S0165-0114(98)00137-7

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International journal of machine learning and cybernetics*, *1*, 43-52. DOI:10.1016/S0165-0114(98)00137-7