



## مدل نظارت شده‌ی ضعیف برای تشخیص مشترک احساس-موضوع از داده‌های متنی

مسعود فاطمی

گرایش هوش مصنوعی و رباتیک / رشته‌ی مهندسی کامپیوتر

دانشکده‌ی برق و کامپیوتر

۹۳۱۴۹۹۴

گزارش پروژه‌ی پایانی درس یادگیری ماشین

دکتر پالهنک

زمستان ۹۴

هدف از تحلیل احساس یا کاوش عقاید استفاده از ابزارهای خودکار برای تشخیص و استخراج اطلاعات مفهومی مانند عقاید، نگرش‌ها، و احساسات بیان شده در متن می‌باشد. در این پروژه یک ساختار احتمالاتی جدید تحت عنوان مدل مشترک احساس-موضوع بر اساس مدل تخصیص دیریکله‌ی پنهان معرفی و پیاده سازی می‌شود، که به صورت همزمان موضوع و احساس را از داده‌های متنی استخراج می‌کند. در کنار مدل‌های نظارت شده که برای تشخیص احساس و طبقه‌بندی متون بر اساس احساس وجود دارد و در بیشتر موارد استفاده از چنین مدل‌هایی در یک دامنه‌ی دیگر منجر به نتایج نامطلوب می‌گردد مدل معرفی شده در اینجا یک مدل نظارت شده‌ی ضعیف می‌باشد که این خاصیت باعث می‌گردد اجرای مدل بر روی دامنه‌های دیگر منجر به تولید خروجی مطلوب گردد و کاهش کارایی در خروجی مدل به وجود نیاید. مدل نهایی پیاده‌سازی شده در اینجا بر روی یک پایگاه داده مورد آزمایش قرار می‌گیرد و نتایج حاصل با توجه به معیار سرگشتی مورد ارزیابی قرار می‌گیرند.

کلمات کلیدی: مدل موضوعی، مدل تخصیص دیریکله‌ی پنهان، مدل مولد، تقریب تغییرات، توزیع پسین

## ۱. مقدمه

در این پروژه هدف تشخیص احساس از داده‌های متنی می‌باشد. مدلی که در اینجا برای هدف مورد نظر پیاده سازی شده است بسط داده شده‌ی ساختار تخصیص دیریکله‌ی پنهان<sup>۱</sup> می‌باشد که از آن برای استخراج همزمان احساس<sup>۲</sup> و موضوع از داده‌های متنی استفاده می‌شود. مدل تخصیص دیریکله‌ی پنهان که در ادامه به صورت کامل شرح داده می‌شود یک مدل موضوعی<sup>۳</sup> می‌باشد. با توجه به تعریفی که برای مدل‌های موضوعی وجود دارد چنین ساختارهایی به عنوان زیر مجموعه‌ای از مدل‌های احتمالاتی مولد<sup>۴</sup> در نظر گرفته می‌شوند. با اجرای یک مدل موضوعی بر روی مجموعه‌ای از اسناد یک چکیده از موضوع‌های موجود در اسنادهای مختلف به عنوان خروجی دریافت می‌شود که بسته به کاربرد مورد نظر می‌توان آن را به عنوان یک نمایش منسجم از مجموعه اسناد در نظر گرفت. یا در کاربردی دیگر این مجموعه موضوع‌های موجود در متن را می‌توان به عنوان یک بردار ویژگی در نظر گرفت و برای مثال در فرآیند طبقه‌بندی<sup>۵</sup> متون از آن استفاده کرد.

همان‌طور که ذکر شد مدلی که در اینجا بسط داده می‌شود گسترش یافته‌ی مدل تخصیص دیریکله‌ی پنهان می‌باشد که علاوه بر موضوع، احساس موجود در یک مجموعه سند که می‌تواند مجموعه‌ای از بازبینی‌ها باشد را برای ما استخراج می‌کند. همان‌طور که در پیشنهاد پروژه که پیش از این اعلام گردیده بود برای پیاده سازی این مدل از نرم افزار متلب استفاده شده و کد نوشته

<sup>1</sup> Latent Dirichlet Allocation (LDA)

<sup>2</sup> Sentiment

<sup>3</sup> Topic Model

<sup>4</sup> Generative Probabilistic Models

<sup>5</sup> Classification

شده در محیط این برنامه قابل اجرا می باشد. در ادامه با خروجی ها، ورودی ها، بخش های مختلف پیاده سازی و نحوه ی اجرا به صورت کامل آشنا می شویم.

## ۲. تخصیص دیریکله ی پنهان

همان طور که پیش از این نیز ذکر شد تخصیص دیریکله ی پنهان یک مدل احتمالاتی بر روی مجموعه ای از اسناد می باشد. فرضی که در این مدل می شود بر پایه ی این است که هر سند متنی را می توان به صورت یک توزیع مخلوط<sup>۱</sup> بر روی موضوع های مختلف که در آن هر موضوع به وسیله ی یک توزیع بر روی کلمه ها مشخص می شود در نظر گرفت [1].

با توجه با توضیحات بالا مدل LDA فرآیند مولد زیر را برای هر سند  $w$  در مجموعه سند  $D$  در نظر می گیرد [1]:

۱. انتخاب  $N$  از یک توزیع پواسن مانند  $f$ .

۲. انتخاب  $\theta$  از یک توزیع دیریکله با بردار  $\alpha$ .

۳. برای هر  $N$  کلمه مانند  $w_n$

(الف) انتخاب یک موضوع  $z_n$  از توزیع چندجمله ای  $\theta$ <sup>۲</sup>.

(ب) انتخاب یک کلمه مانند  $w_n$  از  $p(w_n | z_n, \beta)$  که یک توزیع احتمالی شرطی و مشروط به موضوع  $z_n$  می باشد.

یک متغیر تصادفی دیریکله ی  $k$ -بعدی مانند  $\theta$  با توجه به رابطه ی توزیع دیریکله به فرم زیر می باشد و چگالی آن از این رابطه بدست می آید [1]:

$$p(\theta | \alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1} \quad (1)$$

که در آن پارامت  $\square$  یک بردار  $k$ -بعدی می باشد که در آن هر  $\square_i > 0$ ، و  $\square(x)$  تابع گاما<sup>۳</sup> می باشد. با توجه به پارامترهای محاسبه شده ی  $\alpha$  و  $\beta$  توزیع مشترک مخلوط موضوع ها یا  $\square$ ، یک مجموعه از  $N$  موضوع  $z$  و یک مجموعه از  $N$  کلمه ی  $w$  از رابطه ی زیر محاسبه می شوند [1]:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta) \quad (2)$$

<sup>1</sup> Mixture Distribution

<sup>2</sup> Multinomial

<sup>3</sup> Gamma Function

که در آن  $p(z_n | \cdot)$  به بیانی ساده  $\square_i$  امین موضوع از بردار احتمالاتی  $\square$  می باشد. با انتگرال گیری از رابطه ی ۲ بر روی  $\square$  و محاسبه سیگما نسبت به  $z$  آنچه که حاصل می شود توزیع حاشیه ای<sup>۱</sup> برای یک سند می باشد که فرمی به شکل زیر دارد [1]:

$$p(\mathbf{w} | \alpha, \beta) = \int p(\theta | \alpha) \left( \prod_{n=1}^N \sum_{z_n} p(z_n | \theta) p(w_n | z_n, \beta) \right) d\theta. \quad (۳)$$

در نهایت با ضرب کردن این احتمال های حاشیه ای برای هر سند در یکدیگر احتمال مورد نظر برای یک مجموعه سند خاص را می توان بدست آورد [1]:

$$p(D | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d. \quad (۴)$$

لازم به ذکر می باشد که ماتریس  $\beta$  که در روابط بالا موجود می باشد ماتریسی با ابعاد  $V * K$  می باشد که در آن  $V$  تعداد کلمات متمایز موجود در مجموعه سند یا به عبارت دیگر تعداد کلمه های موجود در لغات نامه و  $K$  تعداد موضوع های نهان<sup>۲</sup> خواسته شده می باشد. با توجه به این تعریف عناصر ماتریس  $\beta$  برابر احتمال وجود هر کلمه در هر موضوع می باشد. برای مثل  $\beta_{ij}$  نشان دهنده ی احتمال مشاهده ی کلمه ی  $i$ ام در موضوع  $j$ ام می باشد.

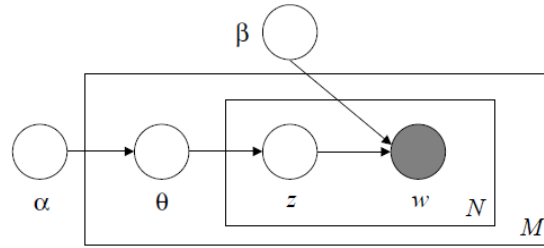
مدل LDA که روابط آن در اینجا شرح داده شد به عنوان یک مدل گرافی احتمالی<sup>۳</sup> در شکل ۱ نشان داده شده است. همان طور که در شکل ۱ نیز مشخص می باشد LDA یک مدل ۳ سطحی می باشد. پارامترهای  $\alpha$  و  $\beta$  پارامترهای در سطح مجموعه سند می باشند و در کل فرآیند تولید یک مجموعه سند تنها یک بار نمونه برداری می شوند و تا انتهای کار ثابت هستند. متغیر  $\theta_d$  یک متغیر در سطح سند می باشد، یا به بیان دیگر برای هر سند یک بار نمونه برداری می گردد. و در انتها متغیرهای  $z_{dn}$  و  $w_{dn}$  متغیرهای در سطح کلمه می باشند که برای هر کلمه در هر سند یک بار نمونه برداری می شوند. لازم به ذکر می باشد که در شکل هایی مانند آنچه که در اینجا به عنوان مدل LDA نشان داده شده است و به آن ها مدل گرافی گفت می شود، پیکان ها نشان دهنده ی وابستگی، مستطیل ها نشان دهنده ی تکرار و دایره ها نشان دهنده ی متغیرهای تصادفی می باشند. البته خود این متغیرهای تصادفی بر دو نوع پنهان (سفید) یا مشاهده شده<sup>۴</sup> (خاکستری) می باشند.

<sup>۱</sup> Marginal Distribution

<sup>۲</sup> Latent Topic

<sup>۳</sup> Probabilistic Graphical Model

<sup>۴</sup> Observed



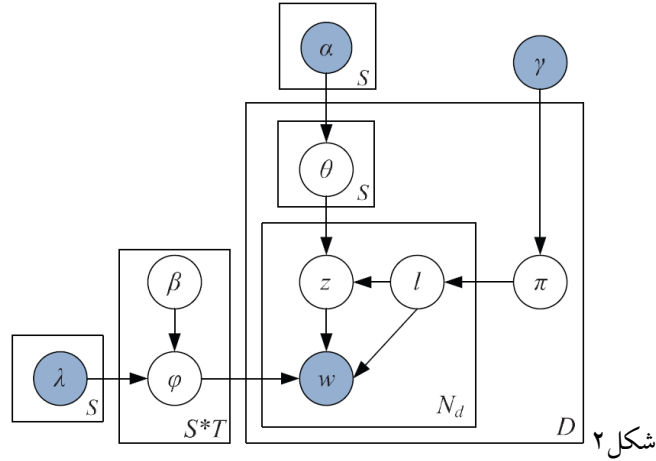
شکل ۱

### ۳. مدل مشترک احساس-موضوع

با توجه به آنچه که در بخش‌های قبل گفت شده مدل مشترک احساس-موضوع<sup>۱</sup> که در شکل ۲ مشاهده می‌شود همانند LDA یک مدل احتمالاتی مولد می‌باشد. در واقع JST گسترش یافته‌ی مدل LDA می‌باشد که علاوه بر تشخیص موضوع به تشخیص احساس از داده‌های متنی نیز می‌پردازد. مدل LDA که در قسمت قبل شرح داده شد بر این فرض می‌باشد که هر سند به صورت مخلوطی از موضوع‌ها می‌باشد، در حالی که هر موضوع یک توزیع احتمالاتی بر روی کلمه‌ها می‌باشد. برای مدل کردن احساس در JST یک لایه بین سند و موضوع اضافه می‌کنیم. با این وجود JST یک مدل چهار لایه می‌باشد که در آن اسناد با احساسات همراه هستند و احساسات با موضوع‌ها همراه هستند و در نهایت کلمات با هر دو برچسب احساس و موضوع همراه هستند.

فرض کنید که یک مجموعه سند در اختیار داریم که شامل  $D$  سند می‌باشد که به صورت  $C = \{d_1, \dots, d_D\}$  مشخص می‌شوند، هر سند موجود دنباله‌ای از  $N_d$  کلمه می‌باشد که با  $d = \{w_1, \dots, w_{N_d}\}$  مشخص می‌شوند و هر کلمه در داخل سند یک مورد از  $V$  کلمه‌ی مجزای واژگان می‌باشد که به صورت  $\{1, \dots, V\}$  شاخص‌گذاری شده‌اند. همچنین  $K$  تعداد برچسب‌های احساس متمایز می‌باشد و  $T$  برابر است با تعداد کلی موضوع‌ها. روند تولید کلمه‌ی  $w_i$  در سند  $d$  تحت JST شامل سه مرحله می‌شود. ابتدا یک برچسب احساس مثلاً  $l$  از  $\pi_d$  که توزیع احتمالاتی احساس برای هر سند می‌باشد انتخاب می‌کنیم. سپس با توجه به برچسب احساس انتخاب شده و مشروط به آن یک موضوع از توزیع موضوعی  $\theta_{d,l}$  انتخاب می‌کنیم. در نهایت یک کلمه از توزیع مشروط به هر دو برچسب موضوعی و احساس برای هر مجموعه سند می‌کشیم و یا تولید می‌کنیم. لازم به ذکر می‌باشد که در مدل LDA به ازای هر سند یک توزیع احتمالاتی برای موضوع‌ها داشتیم، اما در مدل JST به ازای هر برچسب احساس ما یک توزیع احتمالاتی برای موضوع‌ها در سند خواهیم داشت. به طور مثل اگر ما سه برچسب احساس مثبت، منفی یا بی‌طرف داشته باشیم، برای هر سند ما سه توزیع احتمالاتی موضوعی خواهیم داشت که هر کدام متناظر با یک احساس خواهند بود. تفاوت دیگری که بین LDA و JST می‌باشد از این قرار است که در JST به هنگام نمونه‌برداری از کلمه، این کار تنها مشروط به توزیع موضوعی برای مجموعه سند انجام می‌شود ولی در JST این نمونه‌برداری مشروط به هر دو توزیع موضوعی و احساسی انجام می‌شود [2].

<sup>1</sup> Joint Sentiment-Topic (JST)



تعریف رسمی از فرآیند مولد تحت مدل  $JST$  که با شکل ۲ متناظر است به شکل زیر می‌باشد [2]:

۱. برای هر برجسب احساس  $l \in \{1, \dots, S\}$ 
  - برای هر موضوع  $j \in \{1, \dots, T\}$  یک  $\phi_{lj}$  از توزیع  $Dir(\lambda_l * \beta_{lj}^T)$  می‌کشیم.
۲. برای هر سند  $d$  یک توزیع مانند  $\pi_d$  از  $Dir(\gamma)$  بدست می‌آوریم.
۳. برای هر برجسب احساس  $l$  در سند  $d$  یک توزیع مانند  $\theta_{d,l}$  از  $Dir(\alpha)$  انتخاب می‌کنیم.
۴. برای هر کلمه  $w_i$  در سند  $d$ 
  - یک برجسب احساس مانند  $l_i$  از توزیع چند جمله‌ای  $\pi_d$  انتخاب می‌کنیم،
  - یک موضوع مانند  $z_i$  از توزیع چند جمله‌ای  $\theta_{d,l_i}$  انتخاب می‌کنیم،
  - یک کلمه مانند  $w_i$  از  $\phi_{l_i z_i}$  که یک توزیع چندجمله‌ای روی کلمه‌ها و مشروط به موضوع  $z_i$  و برجسب احساس  $l_i$  می‌باشد انتخاب می‌کنیم.

#### ۴. ترکیب کردن دانش اولیه در مدل

همان‌طور که پیش از این نیز ذکر شد مدل  $JST$  گسترش یافته‌ی مدل  $LDA$  می‌باشد. در مقایسه با  $LDA$  در کنار اضافه شدن یک لایه‌ی مولد از احساس همچنین یک وابستگی دیگر از  $\phi$  با توجه به ماتریس  $\lambda$  که اندازه‌ی آن  $S * V$  می‌باشد به مدل  $JST$  اضافه کرده‌ایم که برای ما اطلاعات اولیه‌ی<sup>۱</sup> ما در مورد احساس کلمه‌ها را به صورت کدگذاری شده به سیستم وارد می‌کند. ماتریس  $\lambda$  را می‌توان به عنوان یک ماتریس انتقال<sup>۲</sup> در نظر گرفت که توزیع دیریکله با پارامتر  $\square$  را به گونه‌ای تغییر می‌دهد که می‌توان از آن قطبش<sup>۳</sup> یا جهت احساس اولیه‌ی کلمات را بدست آورد. برای این منظور در ابتدا  $\lambda$  با مقدار ۱ برای تمام عناصر مقدار دهی می‌شود. سپس برای

<sup>۱</sup> Prior Information

<sup>۲</sup> Transform

<sup>۳</sup> Polarity

هر کلمه‌ی  $w \in \{1, \dots, V\}$  در لغات نامه و برای هر برچسب احساس  $l \in \{1, \dots, S\}$  اگر  $w$  در لغات نامه‌ی احساس وجود داشت مقدار  $\lambda_{lw}$  به صورت زیر مقدار دهی می‌شود [2]:

$$\lambda_{lw} = \begin{cases} 1 & \text{if } S(w) = l \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

که در آن  $S(w)$  برچسب احساس اولیه‌ی کلمه‌ی  $w$  در لغات نامه‌ی احساسات را بر می‌گرداند. برای مثل کلمه‌ی "عالی" با شاخص  $i$  در لغت نامه دارای قطبش مثبت از نظر احساسی می‌باشد. در نتیجه بردار سطری متناظر با آن در  $\lambda$  به شکل  $[0, 1, 0]$  خواهد بود. در این حالت برای هر موضوع  $j \in \{1, \dots, T\}$  ضرب  $\lambda_{li}$  با  $\beta_{lji}$  منجر به باقی ماندن تنها  $\beta_{lposji}$  می‌شود و  $\beta_{lnegji}$  و  $\beta_{lneuji}$  برابر با صفر می‌شوند. بنابراین کلمه‌ی "عالی" تنها از توزیع موضوع کلمه‌های مثبت از توزیع دیریکله‌ای با پارامتر  $\beta_{lpos}$  می‌تواند تولید شود.

## 5. استنتاج و تخمین پارامتر

مساله‌ی کلیدی که در بحث استنتاج<sup>1</sup> و به منظور استفاده از مدل نیاز به حل آن داریم محاسبه‌ی توزیع پسین<sup>2</sup> متغیرهای پنهان با توجه به یک سند داده شده می‌باشد [1]:

$$p(\theta, z | \mathbf{w}, \alpha, \beta) = \frac{p(\theta, z, \mathbf{w} | \alpha, \beta)}{p(\mathbf{w} | \alpha, \beta)} \quad (6)$$

متأسفانه این توزیع در حالت کلی قابل حل کردن نمی‌باشد. با توجه به اینکه محاسبه‌ی این توزیع پسین دارای استنتاج مستقیم و دقیق نیست از الگوریتم‌های استنتاجی تغییراتی<sup>3</sup> برای استنتاج پارامترهای مدل استفاده می‌کنند که شامل تقریب لاپلاس<sup>4</sup>، تقریب تغییرات<sup>5</sup> و زنجیره مارکوف مونت کارلو<sup>6</sup> می‌شوند. روشی که در اینجا استفاده می‌شود بر پایه‌ی تقریب تغییرات می‌باشد. ایده‌ی اصلی در استنتاج تغییرات استفاده از نامعادله‌ی جنسن<sup>7</sup> می‌باشد که در آن یک کران پایین قابل تنظیم برای درست نمایی لگاریتم<sup>8</sup> پارامترها بدست می‌آورند. اساساً یک خانواده از کران‌های پایین را در نظر بگیرید که به وسیله‌ی مجموعه‌ای از پارامترهای تغییراتی شاخص گذاری می‌شوند. این پارامترهای تغییراتی با استفاده از یک روش بهینه سازی که قصد پیدا کردن دقیق ترین کران پایین تا حد امکان را دارند انتخاب می‌شوند.

<sup>1</sup> Inference

<sup>2</sup> Posterior Distribution

<sup>3</sup> Variational Inference I

<sup>4</sup> Laplace Approximation

<sup>5</sup> Variational Approximation

<sup>6</sup> Monte Carlo Markov Chain(MCMC)

<sup>7</sup> Jensen's Inequality

<sup>8</sup> Log Likelihood

در بخش تخمین پارامتر از یک روش بیز تجربی<sup>۱</sup> برای تخمین پارامترهای مدل استفاده می‌کنیم. به طور خاص تر با توجه به یک مجموعه سند داده شده به صورت  $D = \{w_1, \dots, w_M\}$ ، ما تمایل به پیدا کردن  $\alpha$  و  $\beta$  به صورتی که داریم که درستنمایی لگاریتمی داده‌ها را برای ما بیشینه کند [1].

$$\ell(\alpha, \beta) = \sum_{d=1}^M \log p(w_d | \alpha, \beta) \quad (7)$$

همان‌طور که پیش از این نیز ذکر شد مقدار  $p(w | \alpha, \beta)$  به صورت عملی قابل محاسبه نمی‌باشد. اگرچه با استفاده از استنتاج تغییراتی می‌توان یک کران پایین عملی برای درست نمایی لگاریتمی پیدا کرد، حال می‌توان این کران را نسبت به  $\alpha$  و  $\beta$  بیشینه کرد. بنابراین می‌توان تخمین بیز تجربی را به صورت تقریبی برای مدل با یک روش متناوب تغییرات حداکثر انتظار<sup>۲</sup> پیدا کنیم. این روش در واقع کران پایین را با توجه به پارامترهای تغییرات بیشینه می‌کند و سپس بعد از همگرایی این قسمت برای یک مقدار ثابت از پارامترهای تغییراتی این کران پایین را نسبت به پارامترهای مدل که  $\alpha$  و  $\beta$  می‌باشند بیشینه می‌کند. روش شرح داده شده را می‌توان به صورت دو مرحله‌ی زیر نوشت:

۱. (E مرحله) برای هر سند مقدار بهینه‌ی پارامترهای تغییرات را پیدا می‌کنیم.
۲. (M مرحله) بیشینه کردن کران پایین حاصل روی لگاریتم درستنمایی با توجه به پارامترهای مدل که  $\alpha$  و  $\beta$  می‌باشند.

## ۶. ارزیابی

برای سنجش میزان توانایی مدل‌های موضوعی در بدست آوردن توزیع کلمات و موضوعی‌ها از معیاری تحت عنوان سرگشتگی<sup>۳</sup> استفاده می‌شود. در اینجا هم از همین معیار برای ارزیابی مدل پیاده‌سازی شده استفاده می‌شود. محاسبه این معیار توسط کد پیاده‌سازی شده که به پیوست تقدیم می‌گردد انجام می‌شود و در خروجی حاصل از اجرای برنامه نمایش داده می‌شود. همچنین در کد پیاده‌سازی شده از این معیار به عنوان شرط همگرایی استفاده شده است و اگر در تکرارهای متوالی مقدار تغییر این معیار از یک حد آستانه کمتر باشد به عنوان شرط خاتمه در نظر گرفته می‌شود و الگوریتم همگرا می‌گردد.

## ۷. نتیجه‌گیری

در این گزارش و پروژه یک مدل نظارت شده‌ی ضعیف برای تشخیص همزمان احساس و موضوع به صورت مشترک از داده‌های متنی را معرفی و پیاده‌سازی کردیم. مدل معرفی شده که JST نام دارد عمل کرد مناسبی برای کاربردهای خواسته شده از خود نشان می‌دهد. خاصیت نظارت شده‌ی ضعیف برای این مدل باعث می‌شود که در مقایسه با سایر مدل‌ها، JST به راحتی قابل انتقال به

<sup>1</sup> Empirical Bayes Method

<sup>2</sup> Variational Expectation Maximization

<sup>3</sup> Perplexity



یک دامنه‌ی دیگر باشد بدون کاهش محسوس در کارایی که در سایر مدل‌ها این اتفاق رخ می‌دهد. از دیگر تفاوت‌های مدل JST با سایر مدل‌ها این است که تمام مدل‌های موجود به تشخیص احساس کلی متن می‌پردازند، این در حالی می‌باشد که JST به تشخیص احساس همراه با هر موضوع و احساس کلی متن می‌پردازد که در مسائل پیش‌رو دید بهتری را به کاربر خواهد داد.

## ۸. پیوست

مدل شرح داده شده در این گزارش در نرم افزار متلب پیاده‌سازی گردیده و فایل‌های آن به پیوست این گزارش تقدیم می‌گردد. پیاده‌سازی صورت گرفته از ۱۱ تابع اصلی تشکیل گردیده که بر روی یک زیر مجموعه‌ی ۱۰۰ تایی از ۱۰۰ بازبینی ابتدایی موجود در پایگاه داده‌ی بازبینی فیلم<sup>۱</sup> آموزش دیده و مقدار سرگشتگی و پارامترهای مدل را محاسبه کرده و به عنوان خروجی بر می‌گرداند. برای اجرای پیاده‌سازی صورت گرفته کفایت دستور زیر در خط فرمان نرم افزار متلب اجرا شود:

```
[alpha,beta,lambda,phi,sentiment_gamma] = ldamain('train',20)
```

## ۹. منابع

- [1] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." the Journal of machine Learning research 3 (2003): 993-1022.
- [2] Lin, Chenghua, et al. "Weakly supervised joint sentiment-topic detection from text." *Knowledge and Data Engineering, IEEE Transactions on* 24.6 (2012): 1134-1145.

---

<sup>1</sup> Movie Review