طبقه‌بندی معنایی تصاویر متنی مبتنی بر الگوی ارزشمندی متن

حسین پورقاسم[۱] - محمد صادق هل فروش[۲] - سبلان دانشور[۳]

(۱) استادیار گروه برق - دانشکده برق - دانشگاه آزاد اسلامی واحد نجف‌آباد

(۲) استادیار گروه برق - دانشکده مهندسی برق و الکترونیک – دانشگاه صنعتی شیراز

(۳) استادیار گروه مهندسی پزشکی – دانشگاه صنعتی سهند

تاریخ دریافت: بهار ۱۳۸۸ تاریخ پذیرش: تابستان ۱۳۸۹

خلاصه: استخراج اطلاعات از تصاویر متنی شناسایی شده از مجرای اینترنت یکی از مهمترین معضلات سیستم‌های جمع‌آوری اطلاعاتی در حوزه تکنولوژی اطلاعات می‌باشد. این مسأله وقتی حادتر می‌گردد که بدانیم در بین انبوه تصاویر متنی تنها درصد ناچیزی از تصاویر متنی شناسایی شده، دارای ارزش اطلاعاتی می‌باشند. در این مقاله، جهت تجزیه و تحلیل تصاویر متنی و دسترسی به محتوای آنها، از یک روش طبقه‌بندی مبتنی بر ناحیه‌بندی تصویر استفاده شده است. در این الگوریتم به کمک یک روش ناحیه‌بندی دومرحله‌ای، نواحی تصویر شناسایی شده، سپس به کمک یک ساختار سلسله‌مراتبی طبقه‌بندی، نوع ناحیه از لحاظ متنی یا عکسی (غیرمتنی) بودن مشخص می‌گردد. در ادامه با تعریف باارزش بودن متن یک تصویر متنی، سعی می‌شود تصویر متنی در یکی از دو گروه معنایی باارزش و بی‌ارزش دسته‌بندی گردد. الگوریتم پیشنهادی بر روی یک پایگاه داده تصاویر متنی و غیرمتنی که از تصاویر موجود در اینترنت فراهم شده است، موردارزیابی قرار می‌گیرد. نتایج آزمایشها بیانگر کارایی روش پیشنهادی در طبقه‌بندی معنایی تصاویر براساس تعریف کاربر از باارزش و بی‌ارزش بودن تصاویر متنی است. الگوریتم پیشنهادی، صحت طبقه‌بندی 98.8% برای طبقه‌بندی تصاویر متنی باارزش از بی‌ارزش را فراهم کرده است.

کلمات کلیدی: طبقه‌بندی محتوایی، تصاویر متنی و غیرمتنی، ارزشمندی اطلاعاتی، سیستم‌های جمع‌آوری اطلاعات.

۱- مقدمه

رشد فزاینده اسناد کاغذی دیجیتالی (اسناد کاغذی اسکن‌شده) در مجاری اینترنتی و سامانه‌های خودکارسازی ادارای، نیاز به بایگانی و دسترسی آنها جهت استخراج و داده‌کاوی اطلاعات را به یک نیاز مهم تبدیل کرده است. با توجه به زمان‌بر بودن این فرآیند به‌صورت دستی (توسط کاربر)، ارائه ابزار خودکاری که بتواند این نیاز را مرتفع کند، ضروری به نظر می‌رسد. بنابراین در یک سامانه خودکار جستجو و بازیابی اطلاعات متنی، تجزیه و تحلیل اسناد متنی[۱] جهت دسترسی به اطلاعات این اسناد امری اجتناب‌ناپذیر تلقی می‌شود [۱]. تجزیه و تحلیل اسناد اسکن‌شده، مقدمه‌ای بر دسترسی به اطلاعات معنایی و مفهومی نهفته در تصویر متنی است. این اطلاعات سطح بالا جهت طبقه‌بندی معنایی تصاویر می‌تواند ابزار نیرومندی در اختیار طراحان سیستم‌های جمع‌آوری اطلاعات قرار دهد. اسناد به سه روش عمده تجزیه و تحلیل می‌شوند[۲]:

۱- تجزیه و تحلیل بر اساس محتویات سند: در این دسته روش صرفنظر از اینکه چطور اسناد ایجاد شده‌اند، براساس محتویاتشان به دو دسته تقسیم می‌شوند: طبیعی (مانند عکس صورت و تصاویر ماهواره‌ای) و نمادین (مانند کارهای چاپی، فرم‌های اداری و طراحی‌های مهندسی) [۲].

۲- تجزیه و تحلیل بر اساس لایه‌بندی سند: در این دسته روش، لایه‌بندی سند مورد بررسی قرار می‌گیرد. این دسته روش نیز به نوبه خود به دو دسته تقسیم می‌شود: تجزیه و تحلیل هندسی (فیزیکی) و تجزیه و تحلیل منطقی. تجزیه و تحلیل هندسی، خصوصیات هندسی است که مناطق همگون مانند متن، جدول، تصویر، نقاشی و ... را دسته‌بندی می‌کند. در مقابل، تجزیه و تحلیل منطقی، به نوع صفحه‌بندی می‌پردازد و برچسب‌های موجود در صفحه مانند عنوان، علامت تجاری، پاورقی، حروف بزرگ و کوچک و ... را مشخص می‌کند [۱،۳].

۳- تجزیه و تحلیل بر اساس متن: در این تجزیه و تحلیل، سند به دو قسمت متنی و غیرمتنی (گرافیکی) تقسیم می‌شود. برای دسته‌بندی و

جداسازی نواحی متنی از نواحی گرافیکی، از فیلترهای موجود در پردازش تصویر استفاده می‌شود [۴- ۵].

اکثر کارهایی که اخیرا انجام شده است، تجزیه و تحلیل سند را بر اساس ناحیه‌بندی تصاویر سندی به نواحی متنی و گرافیکی موجود در تصویر انجام داده‌اند (دسته روش سوم). در این تحقیقات ناحیه‌بندی معمولا به یکی از سه روش بالا به پایین (Top-Down)، پایین به بالا (Bottom-Up) و مبتنی بر بافت (Texture) انجام شده است [۶،۷]. روش‌هایی که به صورت پایین به بالا عمل می‌کنند، از اجزای کوچک تصویر (پیکسل) استفاده کرده و با اتصال این اجزا به‌هم و تشکیل اجزای بزرگتر برای شناسایی قطعات اصلی موجود در تصویر استفاده می‌کنند. به عبارت دیگر این روشها، ناحیه‌بندی را از پیکسل شروع کرده و به ناحیه یا شئی می‌رسند. اشکال این روش در حساس بودن به اندازه حروف، وضوح سند اسکن‌شده و فاصله میان حروف است [۴]. از جمله این روش‌ها می‌توان به الگوریتم Docstrum اشاره کرد که بر اساس K- نزدیکترین همسایگی از اجزا صفحه عمل می‌کند [۸]. در روش‌هایی که به صورت بالا به پایین عمل می‌کنند، عملیات ناحیه‌بندی از بزرگترین جزء، یعنی تمام صفحه شروع شده تا به کوچکترین قطعه در صفحه می‌رسد. اشکال این روش در تشخیص زاویه و انحنای متن است. زیرا محدود به بلوکهای مستطیلی است [۴]. از جمله این روش‌ها می‌توان به الگوریتم بازگشتی X-Y Cut [۹] اشاره کرد. در هر مرحله مقادیر واریانس در جهت‌های افقی و عمودی محاسبه شده و با مقدار آستانه‌ای که از پیش تعریف شده است، مقایسه می‌گردد و در انتها نیز ناحیه‌بندی انجام می‌گیرد. روش‌هایی که بافت را برای ناحیه‌بندی به کارمی‌برند، از ویژگیهای بافت تصویر برای ناحیه‌بندی استفاده می‌کنند [۱۰]. در دو روش اول (پایین به بالا، بالا به پایین) نیاز به پیش اطلاعاتی درباره فاصله‌های افقی و عمودی، مجزا بودن متن‌ها، عکس‌ها، گرافیک‌ها و حتی فرض‌هایی درباره متن، سبک و اندازه می‌باشد. با توجه به اشکالات و موارد مطرح شده در مورد روشهای ناحیه‌بندی، در این مقاله از روش ناحیه‌بندی مبتنی بر بافت استفاده می‌شود.

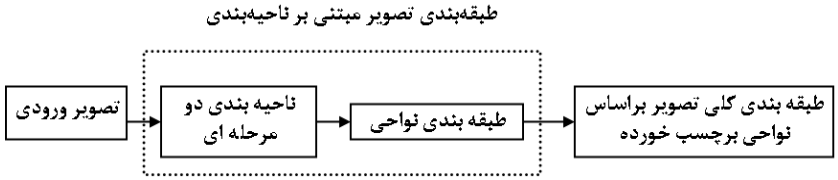
در بسیاری از روشهای ارائه شده قبلی در زمینه طبقه‌بندی، برای آنکه ویژگیها با اطلاعات محلی در اختیار قرار گیرند تا عملکرد طبقه‌بندی بهبود یابد، ویژگیها به صورت محلی استخراج می‌شوند [۱۱]. برای این منظور، تصویر به چندین ناحیه ثابت مثلا بلوکهای مستطیلی هم اندازه تقسیم شده و ویژگیها از هر بلوک استخراج می‌شوند. در این مقاله، با استفاده از یک ساختار طبقه‌بندی سلسله مراتبی مبتنی بر ناحیه‌بندی تصاویر سندی ارائه شده در [۱]، نواحی متنی و گرافیکی تصویر ورودی مشخص می‌شود. در این نقشه، از یک الگوریتم ناحیه‌بندی دو مرحله‌ای مبتنی بر تبدیل موجک و آستانه‌گذاری استفاده می‌شود. در اولین مرحله از الگوریتم ناحیه‌بندی، از تبدیل موجک [۴-۷،۱۲، ۵] برای ناحیه‌بندی استفاده می‌شود. سپس با استفاده از یک طبقه‌بند شبکه عصبی پرسپترون چندلایه (MLP)، برچسب نواحی از نظر متنی بودن یا گرافیکی (غیرمتنی) بودن مشخص می‌گردد و نواحی متنی شناسایی شده از تصویر کنار گذاشته می‌شوند. در مرحله دوم با اعمال یک ناحیه‌بندی ثانویه مبتنی بر آستانه‌گذاری بر روی باقی‌مانده تصویر حاصل از ناحیه‌بندی اولیه، تصویر به صورت کامل ناحیه‌بندی می‌شود. در انتها نیز برچسب نواحی ناحیه‌بندی شده توسط ناحیه‌بند دوم توسط دو طبقه‌بند (MLP و ماشین بردار پشتیبان (SVM)) مشخص می‌گردد. در هر سه طبقه‌بند، از ویژگیهای بافت جهت طبقه‌بندی نواحی استفاده می‌شود. ویژگیهایی همچون Tamura (شامل زبری، جهت‌دار بودن و کنتراست)، همبستگی، انرژی، همگنی و آنتروپی مستخرج از ماتریس همرخدادی و همچنین دو ویژگی از زیرباندهای تبدیل موجک استخراج می‌گردد. در ادامه با تعریف الگوی ارزشمندی متن و استخراج ویژگی از این الگو، یک طبقه‌بند شبکه عصبی آموزش داده می‌شود تا در مرحله آزمایش تصاویر متنی باارزش از بقیه تصاویر مجزا گردد.

در ادامه این مقاله، در بخش (۲) چارچوب پیشنهادی طبقه‌بندی معنایی تصاویر معرفی می‌گردد. در این بخش، جزئیات ناحیه‌بندی دومرحله‌ای، استخراج ویژگی، تعریف الگوی ارزشمندی متن و طبقه‌بندی معنایی تصویر تشریح می‌شود. در بخش (۳) آزمایشهای انجام شده جهت ارزیابی الگوریتم پیشنهادی آورده شده است. بخش (۴) مروری بر عملکرد الگوریتم پیشنهادی ارائه می‌دهد. در انتها نیز در بخش (۵) نتیجه‌گیری از کار انجام شده در این مقاله ارائه می‌گردد.

۲- چارچوب پیشنهادی طبقه‌بندی معنایی تصاویر

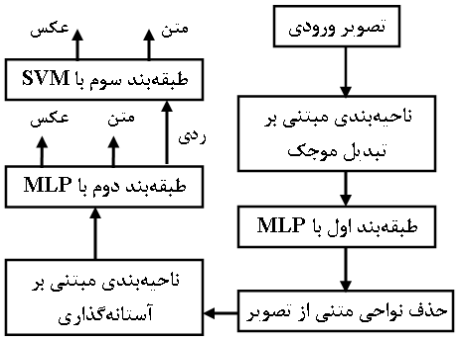
شکل (۱) روندنمای کلی چارچوب پیشنهادی برای طبقه‌بندی معنایی تصاویر متنی مبتنی بر طبقه‌بندی نواحی تصویر را نشان می‌دهد. در چارچوب پیشنهادی تصویر ورودی، طی دو مرحله ناحیه‌بندی، به یک سری نواحی با برچسبهای متنی و گرافیکی ناحیه‌بندی می‌شود. در مرحله آخر نیز، با آموزش الگوهای تصاویر متنی باارزش و بی‌ارزش تصاویر مطلوب و نامطلوب از یکدیگر جدا می‌شوند. در این مقاله جهت طبقه‌بندی نواحی مستخرج از تصویر از الگوریتم معرفی شده در [۱] استفاده شده است. روندنمای طبقه‌بند سلسله‌مراتبی مبتنی بر ناحیه‌بندی با جزئیات بیشتر در شکل (۲) نشان داده شده است.

در [۱]، از یک الگوریتم ناحیه‌بندی دومرحله‌ای برای ناحیه‌بندی تصویر و به دنبال آن طبقه‌بندی نواحی استفاده شده است. در مرحله اول این الگوریتم، تصویر ورودی با یک روش ناحیه‌بندی مبتنی بر تبدیل موجک ناحیه‌بندی می‌شود. سپس این نواحی با طبقه‌بند اول، طبقه‌بندی می‌شود. نواحی متنی شناسایی شده از تصویر ورودی حذف شده و نواحی گرافیکی در تصویر باقی می‌مانند تا دوباره و با دقت بیشتر توسط ناحیه‌بند دوم (ناحیه‌بند مبتنی بر آستانه‌گذاری) ناحیه‌بندی شده و به کمک دو طبقه‌بند (طبقه‌بند دوم و سوم)، متنی یا گرافیکی بودن نواحی مشخص می‌گردد. برای استفاده بهتر از الگوریتم طبقه‌بندی مبتنی بر ناحیه‌بندی ارائه شده در [۱]، مرحله استخراج ویژگی از نواحی را تغییر دادیم. بدین صورت که با اضافه کردن یک ویژگی بافت دیگر (ویژگی Tamura) به بردار ویژگی سعی شده است که بردار ویژگی تقویت شود. در ادامه تنها به تشریح ویژگی جدید Tamura پرداخته می‌شود.



شکل (۱): روندنمای کلی چارچوب پیشنهادی طبقه‌بندی معنایی تصاویر

Fig. (1): The general process for the proposed classifying frames for image meaning



شکل (۲): روند نمای طبقه‌بند سلسله‌مراتبی مبتنی بر ناحیه‌بندی [۱]

Fig. (2): The general process of classifying the sequences based on regioning [1]

۲-۱- استخراج ویژگی

مهمترین مرحله در حل یک مسأله شناسایی الگو، مرحله انتخاب و استخراج ویژگی است. در این مقاله، از یک مجموعه کامل ویژگی بافت جهت طبقه‌بندی نواحی تصاویر در ساختار پیشنهادی [۱] استفاده شده است. این مجموعه شامل ویژگیهای Tamura (شامل زبری، جهت‌دار بودن و کنتراست)، همبستگی، انرژی، همگنی و آنتروپی مستخرج از ماتریس همرخدادی [۱۶] و همچنین دو ویژگی مستخرج از زیرباندهای تبدیل موجک می‌باشد [۱]. در ادامه جزئیات استخراج این ویژگیها آورده می‌شود.

۲-۱-۱- ویژگیهای Tamura

ویژگیهای تامورا، شامل زبری[۲]، کنتراست[۳]، جهت‌دار بودن[۴]، شباهت خطی[۵]، منظم بودن[۶] و سختی[۷] می‌شوند [۱۳]. این ویژگیها به واسطه مطالعات روان‌شناسی بر روی ادراک انسان از بافت طراحی شده است. در این مقاله، سه جزء اول این ویژگی استفاده می‌شود که در اینجا نیز جزئیات آنها آورده شده است.

الف- زبری

زبری معیاری از دانه دانه شدن و زبر شدن بافت است. جهت محاسبه زبری، ابتدا با یک پنجره $2^k \times 2^k$ $(k=0,1,\ldots 5)$ بر روی هر پیکسل (x,y)، فیلتر MA(Moving Average)، به صورت زیر اعمال می‌شود:

$$A_k(x,y)=\sum_{i=x-2^{k-1}}^{x+2^{k+1}-1}\sum_{j=y-2^{k-1}}^{y+2^{k+1}-1}\frac{g(i,j)}{2^{2k}} \qquad (۱)$$

در رابطه بالا $g(i,j)$ شدت روشنایی در پیکسل (i,j) است.

سپس تفاضلات بین جفت $A_k(x,y)$های غیرهمپوشان در جهت‌های افقی و عمودی برای هر پیکسل محاسبه می‌شود:

$$E_{k,h}(x,y)=\left|A_k(x+2^{k-1},y)-A_k(x-2^{k-1},y)\right| \qquad (۲)$$

$$E_{k,v}(x,y)=\left|A_k(x,y+2^{k-1})-A_k(x,y-2^{k-1})\right| \qquad (۳)$$

بعد از این، مقدار kیی که E را دریک جهت حداکثر می‌کند، جهت مشخص کردن اندازه پنجره فیلتر برای هر پیکسل استفاده می‌شود.

$$S_{best}(x,y)=2^k \qquad (۴)$$

در نهایت زبری از میانگین S_{best} کل تصویر محاسبه می‌شود یعنی:

$$F_{crs}=\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}S_{best}(i,j) \qquad (۵)$$

ب- کنتراست

رابطه‌ای که از روی آن کنتراست محاسبه می‌شود، به صورت زیر است:

$$F_{con}=\frac{\sigma}{\alpha_4^{1/4}} \qquad (۶)$$

که $\alpha_4=\mu_4/\sigma^4$ ممان چهارم یا Kurtosis است. μ_4 ممان چهارم حول میانگین و σ^2 واریانس است. این رابطه هم برای کل تصویر و هم برای یک ناحیه از تصویر قابل استفاده است.

ج- جهت‌دار بودن

جهت محاسبه جهت‌دار بودن، تصویر با دو ماتریس 3×3 یعنی M_1 و M_2 کانوالو می‌شود و یک بردار گرادیان در هر پیکسل محاسبه می‌شود.

$$M_1=\begin{bmatrix}-1 & 0 & 1\\ -1 & 0 & 1\\ -1 & 0 & 1\end{bmatrix},\ M_2=\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix} \qquad (۷)$$

دامنه و فاز این بردار به صورت زیر تعریف می‌شود:

$$|\Delta G|=(|\Delta_H|+|\Delta_V|)/2 \qquad (۸)$$

$$\theta=\tan^{-1}(\Delta_V/\Delta_H)+\pi/2 \qquad (۹)$$

که Δ_V, Δ_H تفاضلات کانولوشن در دو جهت افقی و عمودی است.

سپس با عددی کردن θ و شمارش تعداد پیکسلهایی که دامنه، $|\Delta G|$، بزرگتر از یک حد آستانه دارند، یک هیستوگرام از θ و H_D ساخته می‌شود. این هیستوگرام در مورد تصاویر جهتی دارای پیکهای بزرگی است و در مورد تصاویر غیرجهتی به صورت مسطح است. در نهایت از

کل هیستوگرام جهت استخراج یک معیار جهتی براساس بلندی پیکها استفاده می‌شود.

$$F_{dir} = \sum_{p}^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi) \qquad (۱۰)$$

در رابطه بالا P پیک، n_p تعداد پیکها و W_p تعداد بیتهایی است که پیک p را شکل داده است. ϕ_p بیتی است که پیک p بر روی آن قرار گرفته است.

۲-۲- تعریف ارزشمندی متن

در این مقاله، جهت انجام طبقه‌بندی معنایی تصاویر متنی شناسایی شده، از یک ساختار قابل تعریف ارزشمندی متن توسط کاربر استفاده می‌شود. بدین معنی که پس از انجام طبقه‌بندی نواحی استخراج شده از تصویر و مشخص شدن متنی یا عکسی بودن هر ناحیه از تصویر توسط الگوریتم طبقه‌بندی مبتنی بر ناحیه‌بندی، با مقایسه فازی این اطلاعات با الگوی از پیش تعریف شده‌ی ارزشمندی متن، میزان ارزشمند بودن متن مشخص می‌شود. بنابراین در تعریف الگوی ارزشمندی متن و فرآیند مقایسه فازی (به عبارت دیگر طبقه‌بندی معنایی تصویر) آن با یک تصویر، دو نکته نهفته است که در ادامه به آن پرداخته می‌شود. در این مقاله، تصاویر متنی مشابه به نامه‌های رسمی و اداری باارزش تلقی می‌گردد. به عبارت دیگر، متن با ارزش باید حداقل شامل نواحی متنی و گرافیکی زیر باشد:

الف- یک ناحیه گرافیکی در بالای صفحه (در وسط یا سمت راست صفحه) که بیانگر علامت تجاری یا لوگوی یک شرکت یا سازمانی باشد.

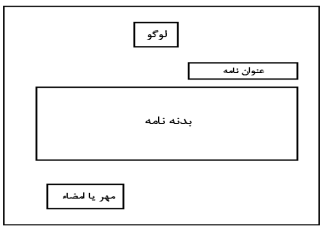
ب- یک ناحیه متنی در زیر لوگو که بیانگر عنوان نامه یا طرفین مکاتبه باشد.

ج- یک ناحیه متنی در زیر ناحیه عنوان که بیانگر بدنه نامه می‌باشد.

د- یک ناحیه گرافیکی در زیر ناحیه بدنه که بیانگر مهر یا امضاء نامه می‌باشد. در موارد ذکر شده فوق دو مسأله به صورت نسبی مطرح می‌گردد. یکی، موقعیت نواحی در صفحه و همچنین موقعیت نسبی نواحی نسبت به یکدیگر است. دیگری، مساحت نسبی هر ناحیه متنی یا گرافیکی می‌باشد. شکل (۳) الگوی نوعی یک متن با ارزش (نامه رسمی- اداری) را نشان می‌دهد.

۲-۳- طبقه‌بندی معنایی تصویر براساس الگوی ارزشمندی

مسأله‌ای که در طبقه‌بندی تصاویر متنی با ارزش و تصاویر متنی بی‌ارزش وجود دارد، بحث تعریف معیار یا به عبارتی الگوی با ارزش بودن متن موجود در تصویر است. همانگونه که مشخص است، تصمیم‌گیری نهایی در مورد محتوای یک تصویر ناشناخته، مسأله‌ای است که بصورت قابل توجهی به نظر و انتظار کاربر از سیستم طبقه‌بند دارد. بدین معنی که با توجه به تعریفی که کاربر از تصویر دلخواه یا غیردلخواه دارد، می‌توان تصمیم‌گیری در این مورد را انجام داد.



شکل (۳): الگوی نوعی یک متن باارزش
Fig. (3): A sample pattern for a valuable text

در بخش قبل الگوی کلی ارزشمند بودن متن تشریح شد. طبق تعریف فوق، ویژگیهایی از الگوی ارزشمندی تعریف و استخراج می‌گردد تا به کمک آن بتوان یک شبکه عصبی پرسپترون چند لایه را آموزش داد. پس از انجام آموزش، از شبکه جهت تعیین مقدار باارزش بودن تصویر ورودی استفاده می‌گردد. در ادامه با یک آستانه‌گذاری ساده بر روی این مقدار فازی می‌توان دسته باارزش بودن (کلاس مطلوب) یا بی‌ارزش بودن (کلاس نامطلوب) تصاویر متنی را تعیین کرد.

ویژگیهایی که از الگوی ارزشمندی استخراج می‌گردد به صورت زیر است:

الف- نسبت مساحت نواحی متنی و عکسی به کل تصویر.

ب- نسبت مساحت نواحی عکسی به نواحی متنی.

ج- مساحت هر ناحیه.

د- موقعیت هر ناحیه (مرکز جرم هر ناحیه).

ه- موقعیت نسبی مراکز جرم هر دو ناحیه (یعنی بالا، پایین، راست، چپ بودن).

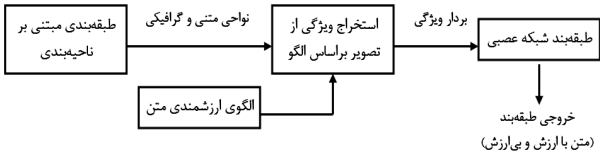
و- نوع ناحیه (متنی یا گرافیکی بودن).

با بردارهای ویژگی شکل گرفته به صورت بالا، جهت آموزش یک شبکه عصبی استفاده می‌گردد. طبقه‌بند مذکور در سه لایه و با الگوریتم پس‌انتشار خطا آموزش می‌بیند تا در مورد نمونه‌های ناشناخته ورودی تصمیم‌گیری کند. شکل (۴) روندنمای طبقه‌بندی معنایی تصاویر متنی مبتنی بر ویژگیهای مستخرج از الگوی ارزشمندی به دو گروه باارزش و بی‌ارزش را نشان می‌دهد.

۳- نتایج آزمایشها

۳-۱- پایگاه داده

در این مقاله، جهت ارزیابی و آزمایش الگوریتم‌های پیشنهادی ناحیه‌بندی و طبقه‌بندی، از یک پایگاه داده تصاویر که از تصاویر موجود در شبکه اینترنت فراهم شده است، استفاده می‌شود. این مجموعه شامل 977 سند تصویری است که به ترتیب شامل 578 سند محض (یعنی تصویری که فقط شامل متن است.)، 81 تصویر محض (تصویر بدون متن) و 318 سند شامل متن و تصویر (مانند روزنامه) می‌باشد. 578 تصویر متنی خالص را می‌توان به اسناد دست‌نوشته، اسناد اداری، کارت‌های تبلیغاتی، روزنامه (چند ستونی)، اسناد تایپی، اسناد جدول‌بندی‌شده، فرم‌های ویندوزی و اینترنتی تقسیم نمود. نمونه‌ای از انواع تصاویر پایگاه داده در شکل (۵) نشان داده شده است.

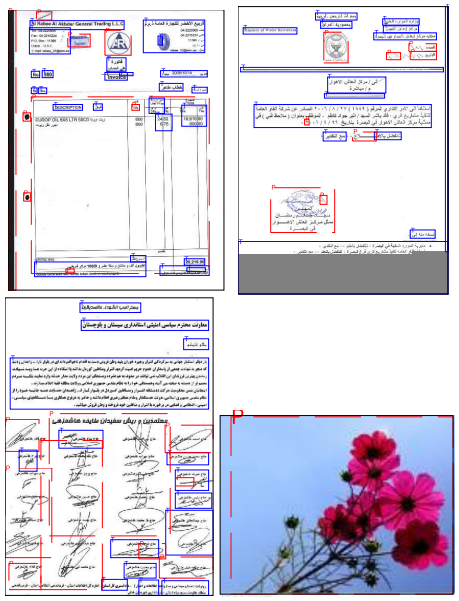


شکل (۴): روند نمای طبقه‌بندی معنایی تصاویر متنی مبتنی بر ویژگیهای مستخرج از الگوی ارزشمندی

Fig. (4): The process of classifying the image meanings, based on the specifications extracted from valuable algorithm

شکل (۵): نمونه‌ای از انواع مختلف تصاویر پایگاه داده.

Fig. (5): Samples of different data based images



شکل (۶): عملکرد الگوریتم پیشنهادی ناحیه‌بندی و طبقه‌بندی بر روی نمونه تصاویر. نواحی متنی با مستطیلهای آبی رنگ با خطوط ممتد و نواحی غیرمتنی با مستطیلهای قرمز رنگ با خطوط خط‌چین مشخص شده‌اند.

Fig. (6): The performance of the proposed algorithm for bloc segmentation and classifying the sample images

۳-۲- طبقه‌بندی نواحی تصویر

الگوریتم طبقه‌بندی مبتنی بر ناحیه‌بندی ارائه شده در [۱]، نواحی مختلف تصویر ورودی را شناسایی کرده و کلاس متنی یا گرافیکی بودن آن را مشخص می‌کند. در این قسمت، نتایج اعمال این الگوریتم بر روی چندین نمونه از تصاویر پایگاه داده نشان داده شده است. نمونه تصاویر بصورتی انتخاب شده است تا اینکه توانایی الگوریتم در حالتها و شرایط مختلف مورد ارزیابی قرار گیرد شکل (۶). در تصاویر ناحیه‌بندی شده شکل (۶) نواحی متنی با مستطیلهای آبی رنگ با خطوط ممتد و نواحی غیرمتنی (نواحی عکسی) با مستطیل‌های قرمز رنگ با خطوط خط‌چین نشان داده شده است. در تصاویر ناحیه‌بندی شده مشاهده می‌گردد که در اکثر موارد نواحی متنی به درستی تشخیص داده شده‌اند و محلهای امضاء، لوگوهای موجود در تصاویر و جداول به عنوان نواحی غیرمتنی (نواحی شکلی) در نظر گرفته شده‌اند. در یکی از تصاویر (گل قرمز) نیز کل تصویر به عنوان ناحیه غیرمتنی در نظر گرفته شده است و هیچ ناحیه متنی شناسایی نشده است. جدول (۱)، نتایج طبقه‌بندی نواحی مستخرج از تمام تصاویر پایگاه داده را نشان می‌دهد. در مجموع طبقه‌بندها 19842 ناحیه را به درستی شناسایی کرده‌اند. در حالیکه 503 ناحیه را نتوانسته‌اند به درستی شناسایی کنند. یعنی در مجموع صحت طبقه‌بندی برابر 97.5% بوده است و تنها 2.5% نواحی شناسایی نشده‌اند جدول (۱).

۳-۳- طبقه‌بندی معنایی تصاویر

جهت بررسی عملکرد طبقه‌بند معنایی پیشنهادی، مجموعه‌ای از تصاویر پایگاه داده با نرم‌افزار طبقه‌بند تصویر به دو گروه تصاویر متنی باارزش (مطلوب) و غیرمتنی (نامطلوب) طبقه‌بندی می‌شود شکل (۷). معیار مطلوب و نامطلوب بودن با توجه به تعریف الگوی ارزشمندی متن در نظر گرفته شده است. به این معنی که تصاویر حاوی نامه‌های رسمی و اداری به عنوان تصاویر با ارزش (تصاویر مطلوب) و بقیه تصاویر (شامل تصاویر windows و web، تصاویر خالص، روزنامه‌های اینترنتی و ...) به عنوان تصاویر بی‌ارزش در نظر گرفته شده است. همانگونه که در شکل (۷) مشاهده می‌گردد، تصاویر خالص که دارای طیف رنگی وسیعی هستند در گروه تصاویر نامطلوب طبقه‌بندی شده‌اند. در مقابل، تصاویر با محتوا رنگی ضعیف در گروه تصاویر مطلوب قرار می‌گیرند. از طرف دیگر، تمام موارد شامل نامه‌های اداری و رسمی، گذرنامه و متنهای روزنامه‌ای به عنوان تصاویر باارزش در نظر گرفته می‌شوند.

Table (1): The accuracy of identified bloc segmentation
جدول (۱): صحت عملکرد طبقه‌بندی نواحی شناسایی‌شده

نوع سند	تعداد موارد صحیح (TA)	تعداد موارد اشتباه (FA)	مجموع	صحت (٪)
دست نوشته	816	23	839	97.25
اداری	1680	38	1718	97.78
کارت‌های تبلیغاتی	142	5	147	96.6
روزنامه	1666	30	1696	98.23
جدول‌بندی شده	2857	75	2932	97.45
فرم‌های مبتنی بر web و windows	4046	93	4139	97.75
شامل متن و تصویر باهم	1970	75	2055	95.86
تصاویر محض	226	39	265	85.28
اسناد دیگر	6439	125	6564	98.09
جمع	19842	503	20345	97.52

تنها موردی که به نظر می‌رسد بی‌ارزش باشد، تصویر مربوط به یک مدار چاپی است که در گروه تصاویر با ارزش دسته‌بندی شده‌اند. تمام موارد شامل تصاویر رنگی، تصاویر فرمهای نرم‌افزارها و صفحات وبگاه‌های اینترنتی به عنوان تصاویر بی‌ارزش در نظر گرفته می‌شوند. در مجموع الگوریتم پیشنهادی، تصاویر متنی باارزش را از بقیه تصاویر با صحت 98.8% طبقه‌بندی می‌کند.

در ادامه برای ارزیابی کامل‌تر الگوریتم پیشنهادی، از یک پایگاه داده با 2390 تصویر اینترنتی استفاده شده است که از نظر فراوانی تصاویر متنی و گرافیکی با دنیای واقعی مطابقت کامل دارد. جدول (۲) نتایج طبقه‌بندی 2390 تصویر اینترنتی را نشان می‌دهد. در این طبقه‌بندی از مجموع 9 تصویر با ارزش تنها یک تصویر نوشتاری با کیفیت پایین در دسته تصاویر بی‌ارزش طبقه‌بندی شده‌است. در مقابل نیز از مجموع 2381 تصویر بی‌ارزش تنها 37 تصویر در دسته تصاویر باارزش طبقه‌بندی شده‌است. به عبارت دیگر درصد پذیرش اشتباه (False Accept) به عنوان تصاویر باارزش 1.5% می‌باشد. این نتایج نشان می‌دهد که کمتر از 2% تصاویر اینترنتی برای بررسی توسط کارشناسان جدا می‌گردد.

Table (2): The results of calssifying the internet images in two valuable images and other images
جدول (۲): نتایج طبقه‌بندی تصاویر اینترنتی به دو گروه تصاویر باارزش و بقیه تصاویر

نوع تصویر	تعداد	نتایج طبقه‌بندی	
		متنی باارزش	بقیه تصاویر
متنی باارزش	9	8	1
بقیه تصاویر	2381	37	2344
مجموع	2390	45	2345

۴- بحث

دسترسی به اطلاعات موجود در تصاویر مسأله مهمی است که همواره مورد نظر محققان زمینه تحقیقاتی پردازش تصویر بوده است. در کاربرد تصاویر متنی، تنها ابزاری که برای نیل به این هدف مورد استفاده قرار می‌گیرد، روشهای تجزیه و تحلیل طرح‌بندی صفحات است. در این دسته روشها، تنها به نوع محتویات و موقعیت فیزیکی آنها پی می‌بریم در حالی که هیچ اطلاعاتی از مفاهیم و معانی نهفته در آنها در اختیار قرار نمی‌گیرد. در این مقاله با در نظر گرفتن این نکته، سعی شد به کمک ویژگیهای سطح پایینی همچون ویژگیهای بینایی تصویر مفاهیم سطح بالای موجود در آن تعریف شود. به عبارت دیگر یک تناظری بین مفاهیم سطح بالا و ویژگیهای سطح پایین برقرار گردد. موفقیت نسبی در برقراری این تناظر حداقل در دو مورد کمک شایانی به سیستم‌های جمع‌آوری اطلاعات می‌کند. اولاً، همانگونه که در بخش (۳-۳) آورده شد. اعمال این چنین الگوریتمی بر روی یک پایگاه داده بزرگ 2390 تصویری، باعث فیلتر شدن قسمت عمده‌ای از تصاویر بی‌ارزش موجود در مجاری اینترنت می‌شود. به عبارت دیگر، به جای اینکه کاربر 2390 تصویر را به صورت دستی مورد بررسی قرار دهد. با این الگوریتم تنها کافی کمتر از 2% آن را یعنی 45 تصویر را مورد بررسی دقیق‌تر قرار دهد. ثانیاً، ایجاد تناظر فوق وابستگی زیادی به نظرات کاربران دارد که این مسأله نکته مثبتی تلقی می‌گردد. چون‌که می‌توان با کمک گرفتن از نظرات کاربران در یک فرآیند بازخورد ربط جهت بهبود تناظر فوق استفاده کرد.

۵- نتیجه‌گیری

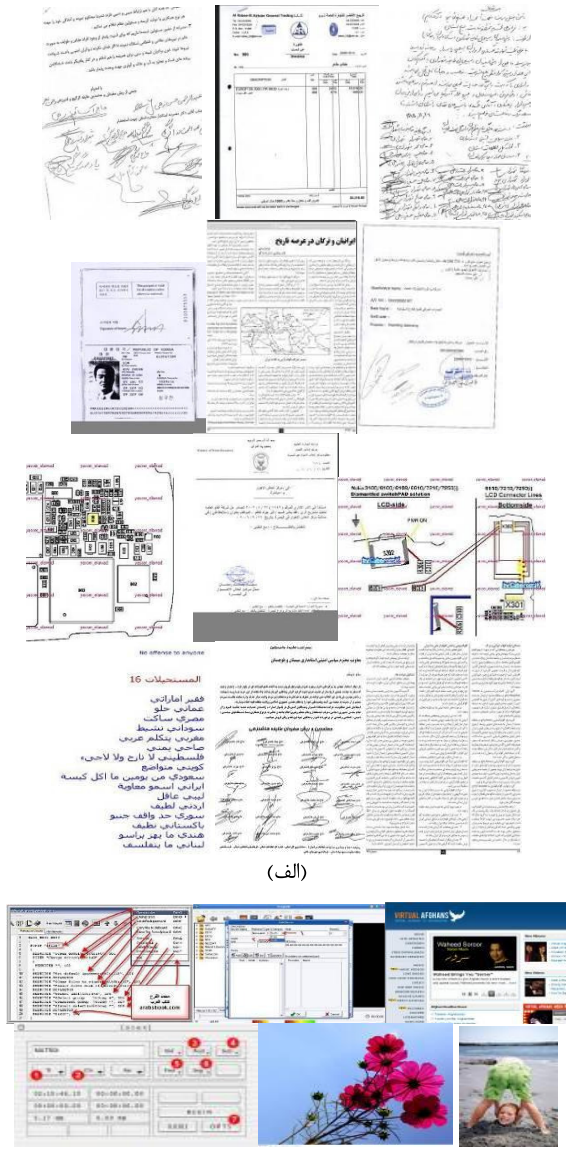
در این مقاله، جهت دستیابی به مفاهیم و معانی موجود در یک تصویر متنی، با استفاده از یک الگوی متنی با ارزش و نتایج حاصل از طبقه‌بندی نواحی تصویر، یک روش کارآمد جهت طبقه‌بندی تصاویر متنی باارزش از بقیه تصاویر ارائه شد. در این ساختار، تصویر به‌وسیله الگوریتم ناحیه‌بندی دومرحله‌ای، ناحیه‌بندی می‌شود. سپس متنی‌بودن و غیرمتنی (عکسی) بودن نواحی حاصل به کمک چندین طبقه‌بند در یک ساختار سلسله مراتبی طبقه‌بندی مشخص می‌گردد. در انتها، با استخراج ویژگی از الگوی متنی باارزش و آموزش یک شبکه عصبی، تصاویر بر اساس محتوای متنی و عکسی، به دو گروه باارزش و بی‌ارزش دسته‌بندی می‌شوند. الگوریتم پیشنهادی بر روی یک پایگاه داده تصاویر متنی و غیرمتنی که از تصاویر موجود در اینترنت فراهم شده است، مورد ارزیابی قرار گرفت. نتایج آزمایشها بیانگر کارایی روش پیشنهادی در طبقه‌بندی معنایی تصاویر براساس تعریف کاربر از باارزش‌بودن و بی‌ارزش‌بودن تصاویر متنی است.

سپاسگزاری

این مقاله مستخرج از طرح پژوهشی مصوب در دانشگاه آزاد اسلامی واحد نجف‌آباد می‌باشد.

پی‌نوشت:

1- Document Image Analysis
2- Coarseness
3- Contrast
4- Directionality
5- Linelikeness
6- Regularity
7- Roughness



(الف)

(ب)

شکل (۷): عملکرد الگوریتم طبقه‌بند پیشنهادی بر روی یک مجموعه تصاویر آزمایشی. (الف) - تصاویر طبقه‌بندی‌شده با برچسب مطلوب و (ب) - تصاویر طبقه‌بندی‌شده با برچسب نامطلوب

Fig. (7): The performance of the proposed bloc segment an algorithm on a collection of experimental images. (a) The classified images with the desired label, (b) The classified images with undesirable labels

References

[1] H. Pour Ghassem, "The analysis of image design based on regional classification in a hieratic structure determination", Jou. of Pow. Elec., Islamic Azad Uni. of Najafabad Branch,No.2, pp.37-44, Spring 2009.

[2] G. Nagy, "Twenty years of document image analysis in PAMI", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.22, pp.38-62, 2000.

[3] N. Amamoto, S. Torigoe, Y. Hirogaki, "Block segmentation and text area extraction of vertically/horizontally written document", CDAR, pp.739-742, Tsukuba, Japan, 1993.

[4] M. Acharyya, M.K. Kundu, "Document image segmentation using wavelet scale-space features", IEEE Trans. Circ. and Sys. for Vid. Tech., Vol.12, No.12, pp.1117-1127, 2002.

[5] A. Busch, W.W. Boles, S. Sridharan, "Texture for script identification", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.27, No.11, pp.1720-1732, Nov. 2005.
[6] S. Mao, T. Kanungo,"Empirical performance evaluation methodology and its application to page segmentation algorithms", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.23, No.3, pp.242-256, March 2001.
[7] Y. Qiao, Z. Lu, C. Song, d.S. Sun,"Document image segmentation using Gabor wavelet and kernel-based methods", Sys. and Cont. in Aer. and Astr., pp.451-455, Jan. 2006.
[8] L. O'Gorman, "The document spectrum for page layout analysis", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.15, pp.1162-1173, 1993.
[9] G. Nagy, S. Seth, M. Viswanathan, "A prototype document image analysis system for technical journals", Computer, Vol.7, pp.10-22, 1992.
[10] A. Jain, Y. Zhong, "Page segmentation using texture analysis", Patt. Rec., Vol.29, pp.743-770, 1996.
[11] J.T. Laaksonen, J.M. Koskela, E. Oja, "Picsom - a framework for content-based image database retrieval using self-organizing maps", SCIA, June 1999, URL: http://www.cis.hut.fi/picsom/publications.html.
[12] Q. Ye, Q. Huang, W. Gao, D. Zhao, "Fast and robust text detection in images and video frames", Ima. and Vis. Comp., Vol.23, No.6, pp.565-576, June 2005.
[13] H. Tamura, S. Mori, T. Yamawaki, "Texture features corresponding to visual perception", IEEE Trans. on Sys., Man., and Cyb., Vol.8, No.6, June 1978.
[14] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, G. Taubin, "The QBIC project: Quering images by content using color, texture, and shape", SPIE, pp.173-187, 1993.
[15] A. Pentland, R.W. Picard, S. Sclaroff, "Photobook: Content-based manipulation of image databases", Int. Jou. of Comp. Visi., Vol.18, No.3, pp.233-254, June 1996.
[16] R.M. Haralick, K. Shanmugan, I. Dinstein, "Textural features for image classification", IEEE Trans. on Sys., Man. and Cyb., Vol.3, No.6, pp.610-621, 1973.
[17] H. Li, D. Doermann, O. Kia, "Automatic text detection and tracking in digital video", IEEE Trans. Ima. Proc., pp.147-156, 2000.
[18] R. Lienhart, A. Wernicke, "Localizing and segmenting text in images and videos", IEEE Trans. Circu. and Sys. for Video Tech., pp.256-268, 2002.

رزومه

حسین پورقاسم استادیار گروه مخابرات دانشکده برق دانشگاه آزاد اسلامی واحد نجف‌آباد، دکتری تخصصی برق با گرایش مهندسی پزشکی را در سال ۱۳۸۷ از دانشگاه تربیت مدرس اخذ کرده است. وی مدرک کارشناسی و کارشناسی‌ارشد خود را در همین رشته در سالهای ۱۳۸۱ و ۱۳۸۳ دریافت کرده است. نامبرده در تمامی این دوره‌ها دانشجوی ممتاز بوده است. زمینه‌های تحقیقاتی مورد علاقه وی شامل زمینه‌های مختلف پردازش تصویر مانند نمایه‌گذاری و بازیابی تصاویر، بیومتریک، ردیابی اشیاء، مخفی‌کردن اطلاعات، شناسایی الگو و شبکه‌های عصبی مصنوعی می‌باشد و در این زمینه‌ها تاکنون بیش از ۲۰ مقاله در مجلات و کنفرانسهای خارجی و داخلی به چاپ رسانیده است.

محمد صادق هل فروش کارشناسی و کارشناسی ارشد خود را در رشته مهندسی الکترونیک به ترتیب از دانشگاه شیراز و دانشگاه صنعتی شریف در سالهای ۱۳۷۲ و ۱۳۷۴ دریافت کرد. او در سال ۱۳۸۵ موفق به اخذ مدرک دکتری در رشته مخابرات گرایش سیستم از دانشگاه تربیت مدرس شد. دکتر هل فروش اکنون استادیار گروه مخابرات در دانشگاه صنعتی شیراز است. زمینه‌های پژوهشی مورد علاقه او شناسایی الگو با تاکید بر سیستمهای بیومتریک و پردازش تصویر است.

سبلان دانشور فارغ‌التحصیل کارشناسی مهندسی پزشکی از دانشگاه علوم پزشکی شهید بهشتی در سال ۱۳۷۵، کارشناسی ارشد مهندسی پزشکی- بیوالکتریک سال ۱۳۸۲ و دکتری مهندسی پزشکی- بیوالکتریک سال ۱۳۸۶ بوده و در حال حاضر استادیار گروه مهندسی پزشکی دانشگاه صنعتی سهند تبریز هستند. زمینه‌های مورد علاقه تحقیقاتی ایشان پردازش تصاویر پزشکی، پردازش سیگنال و مدلسازی بیولوژیکی می‌باشد.

[18] X. Sun, M.H.L. Chow, F.H.F. Leung, D.X.Y. Wang, Y.S. Lee, "Analogue implementation of a neural network controller for UPS inverter applications", IEEE Trans. on Pow. Elec., Vol.17, No.3, pp.305-307, May 2002.

[19] S. Buso, S. Fasolo, L. Malesani, P. Mattavell, "A dead beat adaptive hystersis current control", IEEE Trans. on Indu. Appli., Vol.36, No.4, July/August 2000.

[20] T.S. Lee, S.J. Chiang, J.M. Chang, "H_∞ loop shaping controller designs for the single phase UPS inverters", IEEE Trans. on Pow. Elec, Vol.16, No.4, pp.473-481, July 2001.

[21] G.H. Bode, P.C. Loh, M.J. Newman, D.G. Holmes, "An improved robust predictive current regulation algorithm", IEEE Trans. on Indu. Appl., Vol.41, No.6, pp.1720-1734, Nov/Dec. 2005.

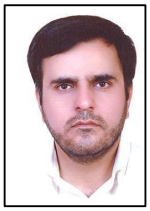
[22] X. Lin, S. Duan, Y. Kang, J. Chen, "Waveform compensation of PWM inverter in UPS", IEEE, Pow. Elect. And Dri. Sys., Proc., Vol.1, pp.292-295, October 2001.

[23] L. Zhous, S.A. Boggs, "High frequency attenuating cable for protection of low voltage ac motors fed by PWM inverters", IEEE Trans. on Pow. Deli., Vol.20, No.2, pp.548-553, April 2005.

[24] R. Zhang, F.C. Lee, D. Boroyevich, H. Mao, "New high power, high performance power converter systems", IEEE Trans. On Powe. Elec., Vol.15, No.3, pp.456-463, May 2000.

Resumeh

Jawad Faiz received his Ph.D. in Electrical Engineering from the University of Newcastle upon Tyne, England in 1988. He is now a Professor at Department of Electrical and Computer Engineering, Faculty of Engineering, University of Tehran. He is the author of 215 publications in international journals and conference proceedings. Dr Faiz is a senior member of IEEE. He is also a member of Iran Academy of Sciences. His teaching and research interests are switched reluctance and VR motors design, design and modeling of electrical machines, drives, and transformers.



Ghazanfar Shahgholian was born in Esfahan, Iran, on December 7, 1968. He graduated in electrical engineering from Esfahan University of Technology (IUT), Esfahan, Iran, in 1992. He received the M.Sc and PhD in electrical engineering from Tabriz University, Tabriz, Iran in 1994 and from Islamic Azad University, Science and Research Branch, Tehran, Iran, in 2006, respectively. He is now an associate professor at Department of Electrical Engineering, Faculty of Engineering, Islamic Azad University – Najafabad Branch. His teaching and research interests include application of control theory to power system dynamics, power electronics and power system simulation.

Mehdi Ehsan received his B.Sc. and M.Sc. in Electrical Engineering from Faculty of Engineering, University of Tehran in 1963, Ph.D. and DIC from Imperial College, University of London in 1976. Since then, he has been with Electrical Engineering Department of Sharif University of Technology in different responsibilities. He has an extended cooperation with research centers in Iran and also abroad. Currently, he is a Professor in Electrical Engineering Department of Sharif University. His research interests include power system simulation, dynamic and transient stability, application of expert systems in identification, operation, planning, and control of power system.