

محاسبه ارزش اعتماد سرویس دهندگان جدید با استفاده از روش داده کاوی KNN

نعیمه ارسطویی، دانشجوی کارشناسی ارشد رشته تجارت الکترونیک، arastooie@aut.ac.ir

امین امیدوار، دانشجوی کارشناسی ارشد رشته تجارت الکترونیک، aminomidvar@aut.ac.ir

حمید رضا شهریاری، عضو هیات علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر،

shahriari@aut.ac.ir

محمد کاظم اکبری فتیله‌ی، عضو هیات علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر،

akbarif@aut.ac.ir

چکیده

با توجه به گسترش شبکه‌ی جهانی اینترنت و همه‌گیر شدن آن در سال‌های اخیر، تجارت الکترونیکی به شکل فزاینده‌ای گسترش یافته و توجه بسیاری از اقشار مختلف مردم را به خود جلب کرده است. به طوری‌که امروزه اغلب سرویس‌دهندگان، سرویس‌های خود را به صورت الکترونیکی به مشتریان ارائه می‌کنند. در کنار هر مبادله‌ی تجاری، بحث اعتماد میان فروشنده و خریدار مطرح می‌شود به گونه‌ای که در صورت عدم وجود اعتماد، تراکنشی نیز انجام نخواهد شد.

در این مقاله مدلی برای محاسبه‌ی اعتماد در فضای کسب و کار الکترونیکی ارائه شده است. مدل مذکور، میزان اعتماد به سرویس‌دهندگان را در زمینه‌های گوناگون با استفاده از فرهنگ لغت wordnet محاسبه می‌کند. همچنین با استفاده از روش‌های داده‌کاوی از جمله KNN، میزان اعتماد به سرویس‌دهندگان جدید را از روی شباهت آن با سایر سرویس‌دهندگان محاسبه می‌نماید. همچنین روشی به منظور تعیین اعتماد به سرویس‌دهندگان با استفاده از شبکه‌ی اجتماعی مبادلات تجاری میان آنان ارائه شده است. در نهایت مدل مذکور بر روی داده‌های انجمن برخط تجارت الکترونیکی Epinions اجرا شده و میزان دقت آن محاسبه شده است.

کلمات کلیدی: مدیریت اعتماد، مشابهت زمینه، Wordnet، k-nearest neighbor.

۱. مقدمه:

اعتماد یک مفهوم اساسی در سیستم‌های پویای تجارت الکترونیک است، به معنای اعتقاد یک عامل (اعتماد کننده) به توانایی و تمایل عاملی دیگر (اعتماد شونده) برای تحویل سرویس مورد توافق در زمینه مشخص و در بازه زمانی معلوم می‌باشد. بکارگیری سیستم‌های مبتنی بر اعتماد سبب تسهیل در کنترل کیفیت خدمات یا محصولات می‌شود و تضمینی برای انجام تراکنش‌ها در محیط‌هایی نظیر اینترنت بحساب می‌آیند [۴]. به دلیل رشد تراکنش‌های برخط و فعالیت‌های تجارت الکترونیک در سال‌های اخیر [۶-۸]، طراحی چارچوب شهرت و اعتماد برای سرویس‌های الکترونیکی، زمینه‌ای است که توجه بسیاری به خود جلب نموده است.

^۱ نویسنده مسئول. تلفن: 09134084990

آدرس پست الکترونیک: arastooie@aut.ac.ir

با توجه به اینکه امروزه اعتماد مبنای بسیاری از فعالیت‌ها در اینترنت محسوب می‌شود و روز به روز نقش آن پررنگ‌تر می‌گردد، مقایسه بین زمینه‌ها، مسئله‌ای است که در برقراری رابطه اعتماد امری ضروری بنظر می‌رسد. روش‌های مختلفی برای اندازه‌گیری مشابهت بین زمینه‌ها در کارهای انجام گرفته وجود دارد [۹] که ما از Similarity wordnet برای اندازه‌گیری مشابهت استفاده نمودیم.

محصولات ارائه شده توسط آن جويا می‌شوند در این حالت افراد باید بتوانند وقت کافی برای جواب‌گویی داشته باشند همچنین باید بتوانند بدون تاثیر گرفتن از محیط و روابط نزدیک مانند، خویشاوندی و دوستی، نظرات خود را ارائه دهند. اما در مقاله‌ی ارائه شده فرض بر این است که در صورت نداشتن شهرت سرویس‌دهنده، از افراد در مورد سرویس‌دهنده و محصولات ارائه شده توسط آن نظر نمی‌خواهیم بلکه از نظرات افراد بر روی محصولات مختلف که در سایت موجود است استفاده می‌نماییم. در این حالت، می‌توان به نظرات افراد نسبت به حالت قبل اعتماد بیشتری داشت، همچنین مزیت دیگر روش ارائه شده در این است که در همه‌ی زمان‌ها نمی‌توان از افراد در مورد سایر افراد نظرخواهی نمود همچنین بدلیل استفاده از اطلاعات موجود در سایت در این روش می‌توان تا حدی از سربار محاسباتی، زمان و هزینه کاست.

همچنین برای اندازه‌گیری مقدار شهرت سرویس‌دهنده‌ای که مقدار شهرت آن را نداریم به‌جای بهره‌گیری از نظرات سایر افراد برای محاسبه‌ی مقدار شهرت، از میزان مشابهت نظرات افراد نظردهنده نسبت به خدمات ارائه شده توسط آن سرویس‌دهنده و نزدیکی نظرات آن‌ها به نظرات سرویس‌دهنده‌ی مورد نظر با بهره‌گیری از توابع ریاضی و الگوریتم k-nearest neighbor در داده کاوی مقدار شهرت را محاسبه خواهیم نمود.

در ادامه پس از بررسی کارهای مرتبط در خصوص اعتماد مبتنی بر زمینه در بخش ۲، مفاهیم، ویژگی‌ها و رابطه‌های مورد نیاز برای محاسبه‌ی ارزش اعتماد و روش پیشنهادی خود را در بخش ۳ معرفی خواهیم کرد و در بخش ۴ به ارزیابی روش خود بر داده‌های سایت Epinions پرداخته و در نهایت نیز نتیجه‌گیری خواهیم داشت.

۲. تحقیقات پیشین:

تا کنون تحقیقات فراوانی برای ارائه‌ی مدل اعتماد در کاربردهای مختلف ارائه شده است. ولی تعداد کمی از آن‌ها توانسته اند از نظرات افراد (داده‌های موجود در سایت)، استفاده‌ی موثر نمایند در واقع استفاده از نظرات افراد بیشتر در سیستم‌های پیشنهاددهنده مورد استفاده قرار گرفته است [۲]. در [۸] چارچوبی برای محاسبه‌ی اعتماد در محیط‌های فراگیر پیشنهاد شده است که در آن فاکتورهای اعتماد، QOS و سیار بودن برای هر خدمت در نظر گرفته شده است. سپس روش‌های کلی بمنظور محاسبه این سه فاکتور ارائه گردیده است. همچنین پروتکل ارتباطی و فرمت‌های استاندارد برای تبادل اطلاعات میان سرویس‌دهندگان و سرویس‌گیرندگان پیشنهاد شده است. یکی از معایب این تحقیق این است که در رابطه با نحوه‌ی محاسبه‌ی اعتماد حرفی زده نشده است.

۳. محاسبه اعتماد:

به‌دلیل مجازی بودن ارتباطات در بازارهای الکترونیکی و عدم اطمینان از کیفیت کالای درخواست شده با توجه به تعداد زیاد سرویس‌دهندگان در چنین محیط‌هایی اعتماد به‌عنوان نقش کلیدی در محاسبات وارد می‌گردد. برای محاسبه‌ی میزان اعتماد یک فرد به فرد دیگر علاوه بر تجربه‌ی شخصی فرد، از تجربه‌ی سایر افراد نیز استفاده می‌گردد [۴].

معادله ۱ چگونگی انجام مراحل تشریح شده را نمایش می‌دهد.

$$Re p_a(o, p)^t = \alpha * SExp_a(o, p)^t + (1 - \alpha) * \frac{\sum_{z \in Z} Re p_a(z, p) * Re c_z(o, p)}{\sum_{z \in Z} Re p_a(z, p)} \quad (1)$$

که مفهوم هر یک از متغیرها به شرح زیر خواهد بود:

$Re p_a(o, p)^t$: میزان شهرت فرد o که توسط عامل a روی محصول p

در زمان t نگه‌داری می‌شود.

در اکثر سیستم‌های اعتماد و شهرت امروزی برای محاسبه‌ی شهرت سرویس‌دهندگان، نظر افراد را در مورد سرویس‌دهنده و

$SEXP_a(o, p)$: میزان شهرت فرد o است که توسط عامل a روی محصول p در زمان t با توجه به تعاملات و تجربیات پیشین a نگه‌داری می‌شود.

$Rec_z(o, p)$: توصیه یا نظری است که فرد z در رابطه با شهرت فرد o روی محصول p دارد.

α : ضریبی در بازه‌ی صفر و یک است و میزان تاثیر تجربه‌ی قبلی و نظرات دیگران را در محاسبه‌ی اعتماد تعیین می‌کند.

حال سعی نمی‌نماییم تا مفهوم زمینه را در محاسبه‌ی ارزش اعتماد وارد نماییم، در این راستا ابتدا به تعریف زمینه می‌پردازیم:

«زمینه، هر نوع اطلاعاتی است که می‌تواند وضعیت یک موجودیت را تعریف کند. این موجودیت می‌تواند یک انسان، مکان، یا هر شیء باشد که در نحوه‌ی تعامل بین کاربر و نرم‌افزار، مؤثر است. این موجودیت می‌تواند شامل خود کاربر یا نرم‌افزار نیز باشد.» [7]

همانطور که در پیش بیان شد ما گاهی به برخی از افراد اعتماد زیادی داریم اما این اعتماد الزاماً اعتماد در تمام زمینه‌ها را در بر نخواهد داشت. مثلاً به یک متخصص کامپیوتر اطمینان کاملی وجود دارد، ولی بدون تجربه و تعامل شخصی نمی‌توان به میزان اطلاعات وی از مسائل پزشکی پی برد. ولی گاهی می‌توان اعتماد به فرد در زمینه‌ی کامپیوتر را به زمینه‌ی قطعات الکترونیکی تعمیم داد. این تعمیم را می‌توان از شباهت نسبی مباحث کامپیوتر با تجهیزات الکترونیکی دانست.

در این مقاله منظور از زمینه، با توجه به مفاهیم مختلفی که ممکن است توسط زمینه پوشش داده شود، مانند: محصولات، سطح کیفیت سرویس و... محصول ارائه شده توسط سرویس-دهنده است. در این راستا اگر میزان شهرت سرویس‌دهنده را در یک محصول مشخص نداشته باشیم، سعی نموده‌ایم تا با وارد نمودن اطلاعات مرتبط با ارزش شهرت محصولات مشابه با محصول مورد نظر با در نظر گرفتن میزان مشابهت بین دو محصول با استفاده از توابع مشابهت موجود در دیکشنری wordnet مانند wordnet similarity distance مشابهت دو محصول را بدست بیاوریم.

در این حالت فرض براین گذاشته شده است که میزان مشابهت با میزان فاصله‌ی دو مفهوم مانند s_1 و s_2 رابطه‌ی مستقیم دارد و به مقادیر $[0-1]$ نرمال شده است، که در رابطه (۲) مشاهده خواهید نمود.

$$\text{Distance}(s_1, s_2) = 1 - \text{Similarity}(s_1, s_2) \quad (2)$$

حال با استفاده از رابطه ۳ میزان شهرت سرویس‌دهنده را با استفاده از سایر محصولات مشابه با محصول مورد نظر در سرویس‌دهنده بدست می‌آوریم.

$$\text{Re } p_a(o, p) = \frac{\sum_{p' \in T} \text{Re } p_a(o, p') \times \text{dist}(p, p')}{\sum_{p' \in T} \text{dist}(p, p')} \quad (3)$$

Dist: نشان‌دهنده‌ی میزان فاصله‌ی مشابهت بین دو محصول مورد نظر p و p' است که با استفاده از روش wordnet محاسبه می‌گردد. T : آرایه‌ای در بردارنده‌ی کل محصولات مشابه با محصول مورد نظر است.

حال برای محاسبه‌ی مقدار شهرت اولیه سرویس‌دهنده براساس مشابهت نظرات افراد مختلف و گرفتن فاصله‌ی اقلیدسی بین نظرات آن‌ها عمل می‌نماییم. در این حالت نظرات افراد روی محصولات مشابه و سرویس‌دهندگان دیگر را با نظرات ارائه شده روی نظرات سرویس‌دهنده‌ای که ارزش شهرت او را نداریم فاصله-ی اقلیدسی گرفته و با توجه به مشخص بودن حوزه‌ی تخصص، زمینه‌ی فعالیت سرویس‌دهنده، حوزه‌ی تخصص یا ارزش شهرت را برای سرویس‌دهنده‌ی جدید محاسبه نماییم.

در این راستا و برای تخمین دقیق و موثرتر از مقدار شهرت برای سرویس‌دهنده‌ی جدید، تنها به نظرات حاصل از یک سرویس-دهنده اکتفا نکرده و با بهره‌گیری از روش k-nearest neighbors در داده‌کاوی k سرویس‌دهنده‌ای را که نظرات ارائه شده‌ی آن‌ها مشابهت بیشتری با نظرات ارائه شده روی محصولات مشابه است را برای محاسبه‌ی ارزش شهرت در نظر می‌گیریم. رابطه‌ی ۴ چگونگی انجام مراحل تشریح شده را نمایش می‌دهد.

$$\text{Re } p(o) = \frac{\sum_{f \in KNN} \frac{1}{\text{dis}(o, f)} * \text{Re } p(f)}{\sum_{f \in KNN} \frac{1}{\text{dis}(o, f)}} \quad (4)$$

Dis(o,f): فاصله‌ی اقلیدسی بین دو فرد نظردهنده‌ی o و f روی محصولات مشابه است.

۴. ارزیابی

برای ارزیابی از مجموعه داده‌های معرفی شده در (Massa2009) استفاده نمودیم. همچنین داده‌های کمکی مورد نیاز به صورت دستی از سایت Epinion جمع‌آوری گردید. تنها داده مورد نیاز server-id سرویس‌دهنده‌ی ارائه‌کننده‌ی محصولات می‌باشد.

برای ساده‌تر شدن محاسبات و اجاره پرس‌وجوها مجموعه داده‌های خود را بر روی پایگاه داده MySQL ذخیره کرده و عملیات لازم را بر روی آن انجام می‌دهیم. در میان داده‌های ۸ سرویس دهنده که محصولات آن‌ها در رده محصولات مشابه الکترونیکی بود را انتخاب نمودیم و از بقیه سرویس‌دهنده‌ها صرف نظر کردیم.

ابتدا با روش پیشنهاد شده شهرت سرویس دهنده‌ها را محاسبه می‌کنیم. سپس فرض می‌کنیم از ۸ سرویس‌دهنده یکی از آن‌ها فعالیت خود را به تازگی آغاز نموده و تنها برخی از افراد از سرویس‌های آن خریداری نموده و استفاده کرده‌اند.

در این جا با استفاده از مجموعه داده‌های مورد نظر، شهرت این سرویس‌دهنده را با توجه به شباهت نظرات افرادی که در رابطه با برخی از خدمات و کالاها این سرویس دهنده نظر داده‌اند با سایر سرویس دهنده‌ها، همانطور که در بخش‌های پیشین نیز شرح داده شد، به دست آوردیم. و این آزمایش را هر بار برای یکی از سرویس دهنده‌ها اجرا نمودیم. نتایج در جدول ۱ قابل ملاحظه است.

جدول ۱: مقایسه روش ارائه شده با روش منت

سرویس	مقدار شهرت	مقدار شهرت	خطا	مقدار شهرت	خطای روش
دهنده	پیشنهادی	واقعی	با	تخمینی	منت
دی	داشته	روشنی	روشنی	روشنی	روشنی

	اطلاعات	منت			
۱	۴,۰۶۳	۳,۶۱۴	۰,۱۷	۳,۶۸۵	۰,۰۷
۲	۴,۳۲۳	۴,۳۶۳	۰,۰۴	۳,۹۴۲	۰,۴۲۱
۳	۴,۵۲۷	۴,۵۷۱	۰,۰۴	۴,۴۶۵	۰,۱۰۶
۴	۴,۳۳۵	۴,۸	۰,۴۶	۴,۳۰۶	۰,۴۹۳
۵	۴,۳۷۱	۴,۵	۰,۱۲	۴,۵۷۸	۰,۰۷۸
۶	۴,۳۱۸	۴,۶۶۶	۰,۳۴	۴,۳۶۹	۰,۲۹۷
۷	۴,۳۶۰	۴,۲۱۴	۰,۱۴	۴,۵۲۵	۰,۳۰۱

همانطور که از جدول ۱ مشخص است، خطای روش ارائه شده در مقایسه با روش قبلی در مورد ۴ سرور از ۷ سرور بسیار بهتر بوده و در سایر موارد نیز عملکرد قابل قبولی دارد.

۵. نتیجه‌گیری

روش ارائه شده در این مقاله در محیط بازارهای الکترونیکی است. در این مقاله با ارائه یک رویکرد جدید در سیستم‌های محاسبه اعتماد و شهرت، مدلی را ارائه دادیم که به بهبودی قابل توجه در عملکرد این سیستم‌ها منجر گردد. مدل ارائه شده با توجه به داده‌های واقعی موجود در سیستم Epinions که در عمل در حال بکارگیری است مورد بررسی و ارزیابی قرار گرفت. با توجه به بالقوه آسیب‌پذیر بودن مدل اعتماد به انواع مختلف حملات و آسیب‌ها، فرض بر این گذاشته شده است که داده‌ها از قبل برای محاسبه‌ی ارزش اعتماد تمیز هستند و عمل فیلترینگ برای کشف داده‌های مرتبط با تبانی و تقلب و... برای عدم ورود داده‌های ناصحیح به سیستم انجام گرفته است اما با این حال جلوی برخی از حملات تا حدی گرفته می‌شود.

ویژگی‌های متمایز این روش نسبت به سایر روش‌های پیشین، علاوه بر وارد نمودن عامل زمینه در محاسبات ارزش اعتماد، که با استفاده از دیکشنری Wordnet انجام می‌گیرد.

- [3] Liu, Jinshan; Issarny, Val'erie. 2004. "Enhanced Reputation Mechanism for Mobile Ad Hoc Networks," *Proceedings of Trust Management: Second International Conference*. France.
- [4] Ruohomaa, Sini; Kutvonen, Lea. 2005, "Trust Management Survey," *proceedings of LNCS*. Vol 3477. pp.77-92.
- [5] Noorian, Zeinab; Ulieru, Mihaela. 2010. "The state of the art in trust and reputation systems: a framework for comparison". *Theoretical and Applied Electronic Commerce Research*. Volume 5, Issue 2. Pp 97-117.
- [6] Sabater, Jordi; Sierra, Carles. 2005 "Review on Computational Trust and Reputation Models". *Kluwer Academic Publisher*.
- [7] Artz, Donovan; Gil, Yolanda. 2007. "A Survey of Trust in Computer Science and the Semantic Web". *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*. Volume 5 Issue 2.
- [8] Mascolo, Cecilia; Capra, Licia; Mcnamara Liam. 2006. "Trust and mobility aware service provision for pervasive computing". *Citeseer*.
- [9] Schickel, Vincent; Faltings, Zuber and Boi. 2007. "OSS: A Semantic Similarity Function based on Hierarchical Ontologies". *In IJCAI*, pages 551-556.

همچنین محاسبه‌ی میزان اعتماد به سرویس‌دهندگان که هیچ مقدار اعتماد اولیه‌ای نسبت به آنان نداریم. با روشی برگرفته از TFIDF، یک بردار ویژگی برای سرویس‌دهندگان استخراج می‌شود و سپس با استفاده از روش K-nearest neighbor، تعداد K سرویس‌دهنده‌ی مشابه با سرویس‌دهنده‌ی مذکور انتخاب می‌شوند. فاصله‌ی میان نظرات سرویس‌دهندگان بر اساس بردار ویژگی و با استفاده از فاصله‌ی اقلیدسی محاسبه می‌شوند. سپس با استفاده از رابطه (۴) مقدار اولیه‌ی اعتماد به آن سرویس‌دهنده محاسبه می‌شود.

مراجع:

- [1] Golbeck, Jennifer Ann; Professor Hendler, James; 2005. "Computing and Applying Trust in Web-based Social Networks". *University of Maryland*. PhD thesis.
- [2] Avesani, Paolo; Massa, Paola; Tiella, Roberto. 2005. "A trust-enhanced recommender system application: Moleskiing". *In SAC '05: Proceeding of ACM Symposium on Applied Computing*. pp 1589-1693.